APPROVAL PAGE FOR GRADUATE THESIS OR PROJECT

GS-13

SUBMITTED IN PARTIAL FULFILLMENT OF REQUIREMENTS FOR DEGREE OF MASTER OF SCIENCE AT CALIFORNIA STATE UNIVERSITY, LOS ANGELES BY

Eugene Shvarts
Candidate

Mathematics
Department

TITLE: **Wavelet-Like Representation of Multivariate Periodic Functions on Scattered Data by Periodic Translation Networks**

APPROVED:

Hrushikesh N. Mhaskar Faculty Member	Signature
Michael Hoffman Faculty Member	Signature
Gerald Beer Faculty Member	Signature
Grant Fraser Department Chairperson	Signature

DATE: **March 5, 2012**

WAVELET-LIKE REPRESENTATION OF MULTIVARIATE PERIODIC FUNCTIONS ON SCATTERED DATA BY PERIODIC TRANSLATION NETWORKS

A Thesis

Presented to

The Faculty of the Department of Mathematics

California State University, Los Angeles

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

By

Eugene Shvarts

March 2012

ACKNOWLEDGMENTS

The progress of my entire academic career has largely been the result of concentrated effort by the most colorful characters in my life, pushing and pushing me to breathe fire into my passions so that I may never be overwhelmed by sloth. Well, completely, anyways. Most of the time. Some of the time.

What I mean to say is, that I thank my mother, Alla Krumer, and grandmother, Mary Blekher, for sticking to tradition and raising me to think critically and take the bitter pill when I have to. I thank my stepfather, Max Krumer, for ushering me into adulthood, and providing me with all the tools I may ever need to succeed. I thank my mentor and academic advisor, Hrushikesh Mhaskar, for teaching me to see mathematics whenever I walk, and, bless his heart, for fathoming that I might somehow one day learn to introspect in any nontrivial way. I thank Frank Filbir for his interest in my work, and his indirect input through Hrushikesh Mhaskar. Moreover, I am indebted to the entire slew of faculty from the Math and Physics departments of Cal State LA who have dealt with my idiosyncrasies, but seen good come of them anyways.

I am grateful to have known as many truly close friends as I already have, and they have pushed the bounds of my creativity and experience in ways I once could not imagine. The sheer mass of good karma that I have been blessed with is nothing short of relativistic, and I can only hope to accrue a fraction as much in return for all the good things that have been done for me. Thank you, everyone, for all that has transpired to bring about this moment.

ABSTRACT

Wavelet-Like Representation of Multivariate Periodic Functions

on Scattered Data by Periodic Translation Networks

By

Eugene Shvarts

This thesis is intended to study and develop the machinery of analysis necessary to perform good approximations of multivariate periodic functions defined using partial information, with which one can retrieve local information about their smoothness. The techniques combine aspects of harmonic analysis, approximation theory, and learning theory. Pioneering work in this field by Mhaskar and his collaborators over the past two decades has led to wavelet–like expansions where the terms are defined using global information such as Fourier coefficients, and yet can reveal local smoothness of a target function. New techniques allow explicit constructions of periodic translation networks with *a priori* performance guarantees, without invoking further optimization routines. The novel contribution of this thesis is such a construction in which the terms are defined using scattered data of the target function, and the local smoothness of the target function is characterized by the local behaviour of the constituent networks.

TABLE OF CONTENTS

Acknowledgments iii

Abstract iv

List of Figures vii

1 Learning Theory and the Approximation Paradigm 1

1.1 The Optimization and Approximation Paradigms 4

2 Wavelet-Like Representation by Multivariate Trigonometric Polynomials 15

2.1 Trigonometric Polynomial Fundamentals 15

2.2 Harmonic Analysis Fundamentals 21

2.3 Low-Pass Filters and the Summability Operator 26

2.4 Multvariate Approximation Results 38

2.5 Classification of Smoothness and Local Smoothness 44

3 Wavelet-Like Representation by Periodic Translation Networks 49

3.1 Discretizing Integrals 52

3.2 Approximation on Scattered Data 58

3.3 PTN Construction and Approximation Properties 69

3.4 Classification of Smoothness and Local Smoothness 78

References 83

LIST OF FIGURES

1.1 The image on the left depicts a rat neuron. Neurons fire when their electric potential exceeds a certain threshold value. The term “neural network” arises in conjunction with the similarity found between the architecture of actual neuron cells, and the topology of an artificial neural network, pictured on the right. The hidden layer is the nonlinearity, or activation function, and represents the analogous dependence on electric potential. The image copyrights are held by Testuya Tatsukawa 2010 and Unikom Center 2010-2012, respectively. 10

1.2 Two curves approximating a univariate real-valued function on nine data points. One is the eighth degree interpolating polynomial, whose oscillations are so wide that they are cut off and appear as near-vertical lines. The other is a lower degree polynomial fit through an unknown regression routine in Apple’s Grapher application, the oscillations of which are significantly damped in comparison. 11

2.1 On the left is a plot of the first thirty Fourier coefficients of $f(x) = \sqrt{|\cos x|}$. On the right is the Daubechies wavelet transform $(\varphi_D^3)_4 * f$, whose peaks detect the cusps at $\pi/2$ and $3\pi/2$. Note the domain in this example is $[0, 2\pi]$, and the plot’s x-axis counts multiples of π. Plots found in [22]. 25

2.2 This low-pass filter is S-smooth for every positive integer S. 28

2.3 Some plots of localized kernels in the univariate case. On the left, notice how the summability kernel for a 1-smooth low-pass filter has oscillations which vanish extremely quickly, unlike the Dirichlet kernel (in green). On the right, the summability kernel for a 0-smooth low-pass filter is plotted in green against that of an infinitely smooth filter in blue, and the graph is shown on $[\pi/4, 3\pi/4]$. 30

2.4 The graphs of $\sqrt{|\cos x|}$ and some partial sums (summing to the 2^{nd} and 6^{th} term) of the Weierstrass nowhere-differentiable function, both functions which lie in the approximation space $B^{\infty}_{\infty,1/2}(\mathbb{T})$. Evidently the local smoothness of these two functions is dramatically different away from the singularities of $\sqrt{|\cos x|}$. 47

CHAPTER 1

Learning Theory and the Approximation Paradigm

This thesis is a survey of intertwined topics in learning theory, harmonic analysis, and approximation theory in which we wish to present concerns and methods from each subject as aspects of one overarching problem with an elegant solution. Let $\mathbb{R}$ denote the real numbers, $\mathbb{C}$ denote the complex numbers, and let the notation $n = 1, 2, \ldots$ choose n as a positive integer. For $q = 1, 2, \ldots$, A^q denotes the set of q-tuples $\{(x_1, x_2, \ldots, x_q) | x_i \in A, i = 1, 2, \ldots, q\}$. Boldface symbols will denote vectors, whose number of components, and the domain in which these components take their values, should be clear from context. If $\mathbf{x}$ has q components, we denote them $\mathbf{x} = (x_1, x_2, \ldots, x_q)$.

- **Learning theory** asks, when one receives from some source, data of the form $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^{n}$, what is the functional relationship which either reproduces $f(\mathbf{x}_k) = \mathbf{y}_k$ exactly, or else to a desired degree of accuracy? The provided data is of course finite, and generally created by a natural source which can sometimes be modelled broadly, but for which theoretically determining the explicit functional relationship is prohibitively difficult. The mathematical difficulties here lie in solving optimization problems or developing approximative models, which we will outline in this chapter.

- **Approximation theory** is the quantitative study of the error inherent in modeling a function using a limited model. The constraints involve the complexity of model used, and the amount of information available - generally the model must conform to some physical principles, and the data may consist of finitely many known values, or expansion coefficients. Hence the data may describe either local or global properties of the function, but in any case we require sufficient data to make a specific conclusion - approximation theory makes this connection between quality and density of data, and consequential accuracy of the model in whatever sense is desired explicit and rigorous, and these ideas are used throughout the thesis.

- **Harmonic analysis** connects the information available about a function in various representations. Fundamental are the Fourier series expansion and Fourier transform, along with the concept of smoothness of a function. These ideas are incorporated when classifying precisely how well approximation models recover behavior of target functions, specifically in regard to how localized such an approximation is. These topics are discussed in Chapter 2.

The classical learning theory problem asks how one should go about determining or constructing a function f which somehow captures the information present in some given data, usually in the form of pairs of vectors, $\{(\mathbf{x}_k, \mathbf{y}_k)\}_{k=1}^{n}$. For example, in [10] , Mhaskar and Khachikyan studied the behavior of flour markets, in which the values of three indices determined the price of flour. This problem of *time series prediction* involved using the values of each index for the past two months to predict the value of each index for the next month – the function which accomplished this hence mapped a six-dimensional real vector to a three-dimensional one. In 1978, data from the Boston housing market was collected and used by Harrison and Rubinfeld in [9] to model the average value that a home in some locality should expect to have. This *regression* problem used thirteen variables representing dimensions including crime and tax rates, accessibility to major business centers, ethnic and financial background, and regional pollution, in order to determine the functional reliance of the final value on each of these dimensions. A key feature of the analysis was eliminating noise introduced in the data, in the form of a dimension known to be statistically uncorrelated with the average value, in this case the "river adjacency" dimension.

Access to accurate, highly scalable techniques for machine learning is critical for modern applications, as databases become ubiquitous and potential learning problems involve high-dimensional spaces. Bornstein and collaborators constructed a database of people, specifying some 115 attributes which could be of use in identifying high-value individuals for US Army applications [2]. This *classification* problem involves selecting between finitely many possibilities (i.e. no threat, possible threat, immediate threat) a value corresponding to some function of the (possibly many) input dimensions, and may involve dimension-reduction techniques.

It is useful to note that in each case, even when the output data is a vector, such a function may be broken down into several individual single-output coordinate functions, and any multiple-dimensional output scheme reduces to a single output. This is because the components of the output vector are not interdependent - they each rely only on the input data. Hence for the remainder of the thesis we consider only data of the form $\{(\mathbf{x}_k, y_k)\}_{k=1}^{n}$, and specifically restrict our attention to $\mathbf{x}_k \in \mathbb{R}^q$

and $y_k \in \mathbb{C}$, for all k, where the dimension q will be given or assumed from context.

1.1 The Optimization and Approximation Paradigms

For the following discussion, the expected value of a probability distribution P will be denoted $E[P]$, as in [25]; the following analysis is found there as well. We will not dwell on the rigors of probability theory here, and present this background for motivational purposes. Technical details may be pursued in [8].

Now, without any hint or assumed structure to inform our analysis beyond the data itself, a natural construction is to assume a joint probability distribution $P(\mathbf{x}, y)$ from which the data is drawn. The conditional probability may be represented in this way via the relationship $P(\mathbf{x}, y) = P(\mathbf{x})P(y|\mathbf{x})$. Foreshadowing the introduction of optimization, we are then interested in the function which minimizes the expected average error

$$I[f] := E\left[(y - f(\mathbf{x}))^2\right] = \int_{\mathbb{R}^q \times \mathbb{C}} (y - f(\mathbf{x}))^2 P(\mathbf{x}, y)\ \mathrm{d}\mathbf{x}\,\mathrm{d}y \quad ,$$

over f in some space $\mathcal{F}$ consisting of functions $\mathbb{R}^q \to \mathbb{C}$. Appropriate choice of $\mathcal{F}$ makes this problem well-posed - in [25], the space of differentiable functions with bounded derivative is taken. The ideal candidate for such a minimizer is the *regression function* $f_0(\mathbf{x})$ which evaluates the expected value of returning y from $P(\mathbf{x}, y)$, given $\mathbf{x}$.

Proposition 1.1. *Let $\mathcal{F}$ be the set of differentiable functions with bounded derivative, $\mathbb{R}^q \to \mathbb{C}$. Let the regression function be defined*

$$f_0(\mathbf{x}) = \int_{\mathbb{C}} yP(\mathbf{x}, y)\,\mathrm{d}y \quad .$$

Then $I[f_0] \leq I[f],\ f \in \mathcal{F}$.

Proof. We proceed as follows: add and subtract the regression function when evaluating the expected average error for $f \in \mathcal{F}$, and obtain

$$\begin{aligned}
I[f] &= \int_{\mathbb{R}^q \times \mathbb{C}} (y - f(\mathbf{x}))^2 P(\mathbf{x}, y)\ \mathrm{d}\mathbf{x}\,\mathrm{d}y \\
&= \int_{\mathbb{R}^q \times \mathbb{C}} (y - f_0(\mathbf{x}) + f_0(\mathbf{x}) - f(\mathbf{x}))^2 P(\mathbf{x}, y)\ \mathrm{d}\mathbf{x}\,\mathrm{d}y \\
&= \int_{\mathbb{R}^q \times \mathbb{C}} P(\mathbf{x}, y)(y - f_0(\mathbf{x}))^2\ \mathrm{d}\mathbf{x}\,\mathrm{d}y + \int_{\mathbb{R}^q \times \mathbb{C}} P(\mathbf{x}, y)(f_0(\mathbf{x}) - f(\mathbf{x}))^2\ \mathrm{d}\mathbf{x}\,\mathrm{d}y \\
&\quad + 2\int_{\mathbb{R}^q \times \mathbb{C}} P(\mathbf{x}, y)(y - f_0(\mathbf{x}))(f_0(\mathbf{x}) - f(\mathbf{x}))\ \mathrm{d}\mathbf{x}\,\mathrm{d}y \quad .
\end{aligned}$$

In the third term, notice that if (using Fubini's theorem) we integrate against y first, the factor $\int_{\mathbb{C}} P(\mathbf{x}, y)(y - f_0(\mathbf{x}))\mathrm{d}y$ is zero by the definition of f_0. The second may be freely integrated against y, and the first is seen to be $I[f_0]$, so we have

$$I[f] = I[f_0] + \int_{\mathbb{R}^q} P(\mathbf{x})(f_0(\mathbf{x}) - f(\mathbf{x}))^2 \mathrm{d}\mathbf{x} \geq I[f_0] \quad ,$$

as the integral expression is non-negative, and we are done. Note that f_0 is then unique, as the condition for the integral expression to vanish nontrivially is $f = f_0$. □

This regression function will now be called the *target function*, as we will spend the remainder of the discussion determining how we may best characterize and approximate it. Note from the proof of Proposition 1.1 that the expected average error expression has the decomposition $I[f] = E\left[(f_0(\mathbf{x}) - f(\mathbf{x}))^2\right] + I[f_0]$. In learning theory these terms are denoted *bias* and *variance* respectively; in approximation theory they are *model-based* or *approximation error*, and *estimation error* or *noise*. Our focus lies entirely with the bias term, as we would like to fix an adequate choice of models, and restrict our attention to the behavior of those models. Because the probability distribution itself is unknown, we must be careful to make our estimations *only in the supremum norm.* This point is critical and will be harped on several times in Chapter 3. Whatever inherent estimation error this produces will be considered given, as the focus of this thesis will not be on noise reduction or fine-tuning models. The opposite is true - we would like to achieve the maximum generality possible for a sufficiently broad choice of models, which in our case will involve certain well-known function spaces $\mathbb{R}^q \to \mathbb{C}$. There is an inherent difficulty to achieving arbitrarily accurate approximation in terms of minimizing the expected average error. Enlisting more complicated models allows us to shrink the bias, but will necessarily complicate the expression for, and conditions on the variance - this fact is known in this context as the *bias–variance tradeoff* (discussed in [25]).

One interpretation of the learning theory problem is function extension - considering the given data to be training data for some function f such that $f(\mathbf{x}_k) = y_k$ for all k, we would like to extend whatever dependence is present on this set to a larger domain in $\mathbb{R}^q$. This yields a solution with an *interpolatory condition.* However, the function space containing a potential choice of model will necessarily be quite large in order to reliably interpolate data that may both consist of many points, and have no prior assumed or known distribution – so-called scattered data. Hence we must develop some salient criterion, whether due to parsimony, aesthetics, or physical limitations, by which to select an appropriate model. The traditional method is then to develop a corresponding energy functional, and represent this criterion as minimizing the energy functional - this is called the *optimization paradigm.*

To generalize slightly, let X be a normed linear space, and q a positive integer. We will refer to elements of X as models, and require that each $f \in X$ maps $\mathbb{R}^q \to \mathbb{C}$. Then, a choice of models is represented by $\mathcal{H} \subset X$. In this light, the optimization problem with interpolatory conditions asks, given the energy functional $\|\cdot\|_E : X \to \mathbb{C}$, determine the existence and uniqueness of

$$\arg\min_{f \in \mathcal{H}} \{\|f\|_E : f(\mathbf{x}_k) = y_k \text{ for } k = 1, 2, \ldots, n\} \quad .$$

For instance, in the univariate case, consider the subspace of real-valued functions with continuous second derivative, among continuous real-valued functions. For a given target function, the optimization problem of minimizing the curvature together with interpolatory conditions is solved by a cubic spline function [20]. Polynomial interpolation is of course possible as well, but this situation does not generalize in a straightforward manner to the case of multiple dimensions. To see this, consider $\mathbb{R}^2$. Given data on two points $\mathbf{x}_1, \mathbf{x}_2$, when $y_1 \neq y_2$, no constant map will interpolate the data, but there is not enough information to uniquely determine an interpolating linear map, which has three parameters. Even when given data on three points $\{\mathbf{x}_k\}_{k=1}^3$, should they be collinear, then a linear interpolatory polynomial will exist only when f is already linear, and in that case will still not be unique [1].

Given a differential operator $\mathcal{L}$ acting on functions over $\mathbb{R}^q$, consider its Green's function G. Golomb-Weinberger theory (see [12]) explains that the optimization problem of minimizing the $L^2(\mathbb{R}^q)$ norm of any function f for which $\mathcal{L}f$ exists, with interpolatory conditions, is solved by a linear combination of terms of the form $\{G(\circ, \mathbf{x}_k)\}_{k=1}^n$. When $\mathcal{L}$ is translation invariant, so that $\mathcal{L}f(\mathbf{x} + \mathbf{a}) = \mathcal{L}f(\mathbf{x})$ for $\mathbf{a} \in \mathbb{R}^q$, then the

Green's function take the form $G(\mathbf{x}, \mathbf{s}) = G(\mathbf{x} - \mathbf{s})$, and so the solution to the optimization problem is a network of translates of G. The word *network* here refers to a linear combination taken with respect to different evaluations of an *activation function*. Given a finite set $\mathcal{C}$, common examples of networks include the following:

- A translation network has the form $\mathbf{x} \mapsto \sum_{\mathbf{y} \in \mathcal{C}} a_{\mathbf{y}} \Phi(\mathbf{x} - \mathbf{y})$, with activation function $\Phi : \mathbb{R}^q \to \mathbb{C}$ and parameters $a_{\mathbf{y}} \in \mathbb{C}$.

- A neural network has the form $\mathbf{x} \mapsto \sum_{\mathbf{y} \in \mathcal{C}} a_{\mathbf{y}} \phi(\mathbf{x} \cdot \mathbf{y} - b_{\mathbf{y}})$, with activation function $\phi : \mathbb{R} \to \mathbb{C}$ and parameters $a_{\mathbf{y}}, b_{\mathbf{y}} \in \mathbb{C}$.

- A radial basis function (RBF) network has the form $\mathbf{x} \mapsto \sum_{\mathbf{y} \in \mathcal{C}} a_{\mathbf{y}} \phi(\|\mathbf{x} - \mathbf{y}\|)$, with activation function $\phi : \mathbb{R} \to \mathbb{C}$ and parameters $a_{\mathbf{y}} \in \mathbb{C}$.

The last case, of RBFs, is of particular interest because they appear as solutions for many repeatedly encountered optimization problems of the Golomb-Weinberger variety. Dyn mentions several in [5], one of which we reproduce.

Let ∂_j denote the partial derivative with respect to the j^{th} coordinate x_j. We will use the notation $D^{\mathbf{r}}$ to represent the differential operator $\partial_{x_1}^{r_1} \partial_{x_2}^{r_2} \ldots \partial_{x_q}^{r_q}$. Further, it will be useful to denote the collection of all integer lattice points whose coordinates are non-negative as $\mathbb{Z}_+^q = \{\mathbf{k} \in \mathbb{Z}^q : k_j \geq 0, j = 1, 2, \ldots, q\}$. Analogously, $\mathbb{R}_+^q = \{\mathbf{r} \in R^q : r_j \geq 0, j = 1, 2, \ldots, q\}$. Generally, for any real-valued expression r, we denote r_+ to be r when $r \geq 0$ and 0 otherwise.

EXAMPLE 1.1.1. The *surface splines* are

$$\phi(t) = \begin{cases} t^{2k-q} \log t \ , & q \text{ even} \ , \\ t^{2k-q}, & q \text{ odd} \ , \end{cases} \tag{1.1.1}$$

with k an integer satisfying $2k > q$. The corresponding RBF interpolant minimizes the functional

$$\mathcal{R}_k(f) = \int_{\mathbb{R}^q} \sum_{\mathbf{r} \in \mathbb{Z}_+^q, |\mathbf{r}|_1 = k} (D^{\mathbf{r}} f)^2 \, \mathrm{d}\mathbf{x} \ .$$

Note that $|\mathbf{r}|_1 = |r_1| + |r_2| + \ldots + |r_q|$ is the ℓ^1 norm, which will be introduced formally at the start of Chapter 2. The functions (1.1.1) are fundamental solutions of the k^{th} iterated Laplacian: $\Delta^k \phi(\|\mathbf{x}\|) = c\delta(\mathbf{x})$. □

A benefit of using the combination of optimization and interpolatory conditions for selecting a model is that the data often does arise from some physical process, which may directly inform the appropriate choice of energy functional. Additionally, these conditions (again, especially for physical problems) are prolific in the mathematical literature, and for the solution of many given optimizations, there exist effective algorithms, numerical methods, and theoretical results which simplify matters considerably. However there is a flip-side, which is that real-world data is inherently noisy to some degree, and any type of noise present in data will significantly erode the quality of an approximation by a model satisfying interpolatory conditions.

Should the form of the noise be understood, we may take a loss functional $\|\cdot\|_L : X \to \mathbb{C}$ to represent a factor which damps this noise. Then we need a method for determining the accuracy of the estimation, as we are no longer interpolating. In general, *regularization* is balancing between these two factors. Traditionally, the

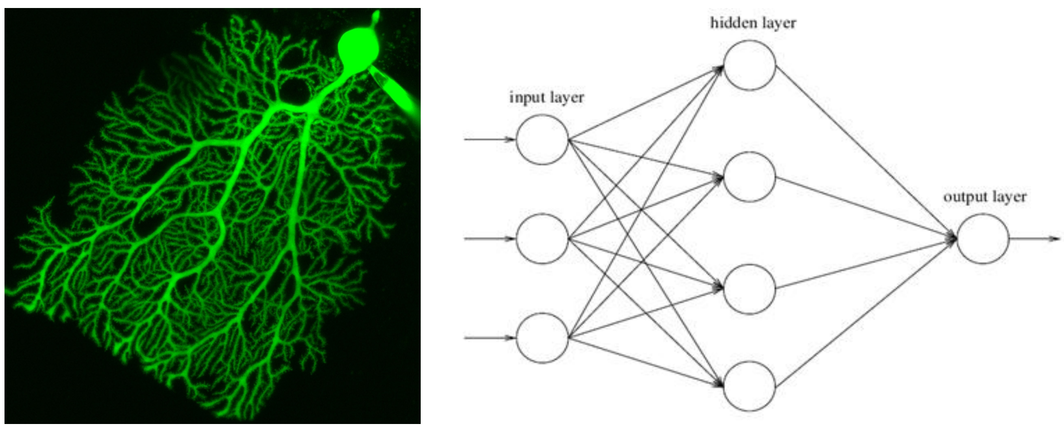

Figure 1.1: The image on the left depicts a rat neuron. Neurons fire when their electric potential exceeds a certain threshold value. The term "neural network" arises in conjunction with the similarity found between the architecture of actual neuron cells, and the topology of an artificial neural network, pictured on the right. The hidden layer is the nonlinearity, or activation function, and represents the analogous dependence on electric potential. The image copyrights are held by Testuya Tatsukawa 2010 and Unikom Center 2010-2012, respectively.

optimization problem with Tikhonov regularization asks, given the loss functional $\|\cdot\|_L$, and a choice of a positive constant δ, determine the existence and uniqueness of

$$\arg\min_{f\in\mathcal{H}}\left\{\sum_{k=1}^{n}(f(\mathbf{x}_k)-y_k)^2+\delta\|f\|_L\right\} \quad .$$

Additionally, we wish to explore the dependence of this quantity on δ. As $\delta \to 0$, we recover the case of pure interpolation. For large δ, the loss functional dominates. This may not be desirable, as for example if we wish to minimize the curvature, taking $\delta \to \infty$ we will recover a linear approximation, which is rarely sufficient. Hence finding the appropriate δ is a matter of fine-tuning to the conditions surrounding each particular problem.

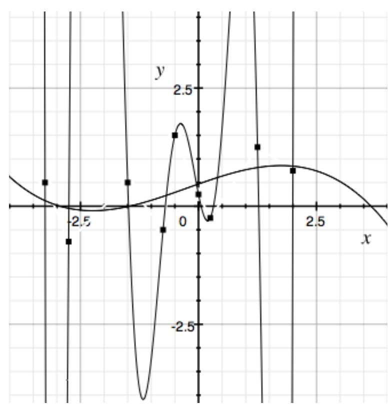

Figure 1.2: Two curves approximating a univariate real-valued function on nine data points. One is the eighth degree interpolating polynomial, whose oscillations are so wide that they are cut off and appear as near-vertical lines. The other is a lower degree polynomial fit through an unknown regression routine in Apple's Grapher application, the oscillations of which are significantly damped in comparison.

Suppose, however, that we wish to determine the functional relationship in some extremely large data set by first solving the corresponding optimization problem on a subset of that data, and then scaling up a seemingly correct solution. This turns out to be impractical, as when the data set changes, the entire optimization problem must be recalculated from scratch. Further, in addition to not informing us what the bounds on the loss functional should be when regularizing, the method does not produce a bound on the size of the minimal function it produces. Given a desired degree of accuracy ϵ, there is no performance guarantee which states when a chosen network is sufficiently large to achieve this accuracy. Most precipitously, by the nature of computational methods for optimization such as steepest descent, we cannot guarantee that solutions found are global optimizers – they may only be local optimizers.

For these reasons, another paradigm is required - instead of imposing conditions on our model in the form of energy and loss functionals, we will base our analysis directly on the target function and systematically approximate it. This is the *approximation paradigm*. Inherently, we now have performance guarantees and bounds on the functionals involved, but along with these come also scalability, broader application, and in fact good estimation to a solution of the optimization problem itself. Our setting is as follows (and as is detailed in [15, 20]):

Let $\mathcal{X}$ be a normed linear space with norm $\|\cdot\|$, $\mathcal{H} \subset \mathcal{X}$, and $f \in \mathcal{X}$. The *degree of approximation* of f from $\mathcal{H}$ is the distance (in the sense of $\mathcal{X}$) between f and $\mathcal{H}$, precisely defined as

$$\mathsf{dist}\ (\mathcal{X}; f, \mathcal{H}) = \inf_{T \in \mathcal{H}} \|f - T\| \quad . \tag{1.1.2}$$

Clearly f is in the closure of $\mathcal{H}$ if and only if $\mathsf{dist}\ (\mathcal{X}; f, \mathcal{H}) = 0$, and we note that considering different norms (equivalent to choosing a different space $\mathcal{X}$) may change this distance as well. The typical situation presents a sequence of nested subsets $\mathcal{V}_0 \subset \mathcal{V}_1 \subset \mathcal{V}_2 \subset \ldots$ of $\mathcal{X}$ - these sets represent our model, and their being nested implies larger indices cannot give larger degrees of approximation. Then we think of $\mathsf{dist}\ (\mathcal{X}; f, \mathcal{V}_m)$ as a nonincreasing function of m, and consider the *rate of approximation* as $m \to \infty$. The natural questions of approximation theory which arise from this context can be summarized as follows:

- Does $\mathsf{dist}\ (\mathcal{X}; f, \mathcal{V}_m) \to 0$ as $m \to \infty$? Judicious choices of $\{\mathcal{V}_m\}$ allows representation of a target function arbitrary well for sufficiently large index (for example the degree of a polynomial, or the number of nodes in a network). This is called the **density** property, because it implies that $\bigcup_{m=0}^{\infty} \mathcal{V}_m$ is dense in $\mathcal{X}$.

- What are the properties of elements v_m^* of $\mathcal{V}_m$ which give the best possible approximation, i.e. $\|f - v_m^*\| = \mathsf{dist}\ (\mathcal{X}; f, \mathcal{V}_m)$? When considering infinite dimensional spaces, when do such v_m^*'s exist? Here the properties of the model are critical, as relying on a best approximation means the resulting solution can inherit undesired behavior. In the case of polynomial approximation for example, the best approximation is characterized by the Chebyshev Alternation Theorem (see [29]). For applications which prefer or require monotone convergence to monotone target functions, best polynomial approximation is then a dead end.

- How are desirable properties of the target function f tied with desirable constructions of the $\mathcal{V}_m$'s? When considering a particular model, *direct theorems* establish expected rates of approximation for a given target function, while *converse theorems* determine properties of a target function based on a given rate of approximation. These can be used to establish limitations on algorithms, and inform their construction. Such conditions are generally posed in terms of smoothness of the function (appropriately defined), or asymptotic behavior of the rate of approximation.

- How does one apply the theory? For instance, when given a data set representing finitely many (potentially noisy) inputs and corresponding outputs to a target function f, how does one construct $v_m(f) \in \mathcal{V}_m$ to be a 'good' approximation as mentioned above? An example of a 'good' approximation condition is

$$\|f - v_m(f)\| \leq c \ \mathsf{dist}\ (\mathcal{X}; f, \mathcal{V}_{\lfloor m/2 \rfloor}) \ ,$$

  with c a constant independent of m and f. The flexibility of this order-of-magnitude bound may reduce computation time in applications. In the process, we would like v_m to have other desirable properties, such as being a linear operator $\mathcal{X} \to \mathcal{V}_m$.

Classically, the most well-known approaches to some of these questions lie in trigonometric polynomial approximation on the unit circle. Exploring these results requires the introduction of the usual associated concepts and function spaces, as discussed in detail in [15] . This will be the primary focus of Chapter 2.

CHAPTER 2

Wavelet-Like Representation by Multivariate Trigonometric Polynomials

In this chapter, we start by introducing the tools of harmonic analysis of periodic functions necessary for our study. Then we describe the development of linear operators which reproduce trigonometric polynomials and provide good approximations to multivariate periodic functions. We conclude with the construction of a function expansion exhibiting wavelet-like properties, permitting analysis of local smoothness.

2.1 Trigonometric Polynomial Fundamentals

Let q be a positive integer, which will be considered a fixed parameter for the remainder of this thesis. The natural domain for periodic complex-valued functions of a single variable is $[-\pi, \pi]$ with the endpoints identified; e.g. for any f, $f(-\pi) = f(\pi)$. We will denote this domain $\mathbb{T}$, and its Cartesian product with itself q times as $\mathbb{T}^q$.

The notion of *measurability* allows us to define norms for functions on $\mathbb{T}^q$. For this chapter we will require only the standard Lebesgue measure, and so for the remainder we will say "measurable" to mean "Lebesgue-measurable". In Chapter 3, we will introduce and define other abstract measures as well, including finitely supported measures when discussing quadrature formulas and the representation of scattered data.

In topology, one defines a continuous function as a map in which the pre-images of open sets are always open. Likewise, a measurable function is a map in which the pre-images of open sets are measurable. We note that the space of measurable complex-valued functions is adequately large for most practical applications of analysis; all Riemann integrable functions are measurable, for instance (see [26]). The Lebesgue measure on $\mathbb{T}^q$, normalized to be a probability measure, will be denoted μ_q^* when referenced explicitly, and the Lebesgue integral of some integrable function f can then be written

$$\int_{\mathbb{T}^q} f(\mathbf{x}) \mathrm{d}\mu_q^*(\mathbf{x}) = \frac{1}{(2\pi)^q} \int_{\mathbb{T}^q} f(\mathbf{x}) \mathrm{d}\mathbf{x} \quad ,$$

where the right-hand side takes the integral with respect to the usual Lebesgue measure on $\mathbb{R}^q$. This is done to simplify the notation.

Now we are in a position to define a notion of size and distance for functions on $\mathbb{T}^q$ - the L^p norms. Let $A \subset \mathbb{T}^q$ and $f : A \to \mathbb{C}$ be measurable. Let $p \in [1, \infty]$. Then the expression

$$\|f\|_{p,A} := \begin{cases} \left\{ \int_A |f(\mathbf{x})|^p \mathrm{d}\mu_q^*(\mathbf{x}) \right\}^{1/p}, & \text{if } 1 \le p < \infty \quad , \\ \operatorname*{ess\,sup}_{\mathbf{x} \in A} |f(\mathbf{x})|, & \text{if } p = \infty. \end{cases} \tag{2.1.1}$$

defines for each p a seminorm on the space of measurable functions $A \to \mathbb{C}$, which is a norm that sends functions other than the constant zero map to zero. The functions which this norm cannot distinguish from the zero map are those which are zero almost everywhere. Precisely, $f, g : A \to \mathbb{C}$ are equal *almost everywhere* (a.e.) when $\mu_q^*(\{\mathbf{x} | f(\mathbf{x}) \neq g(\mathbf{x})\}) = 0$. To circumvent this difficulty, let $f \sim g$ when $f = g$ a.e.; then $\sim$ is an equivalence relation. Note that ess sup(f), the essential supremum, can then be defined as the infimum over the suprema of all functions equal to f a.e..

Precisely,

$$\operatorname*{ess\,sup}_{\mathbf{x}\in A} f(\mathbf{x}) = \inf_{g\sim f}\left\{\sup_{\mathbf{x}\in A}\{g(\mathbf{x})\}\right\} \quad .$$

For the remainder of the discussion, we are not interested in functions that differ on a set of measure zero, so we will join the majority of mathematicians in abusing notation and use f to refer to the equivalence class of functions equal to f almost everywhere. When f is equal a.e. to a continuous function, then that function will represent its equivalence class.

For each p, (2.1.1) defines a norm on the space of all equivalence classes of measurable functions $A \to \mathbb{C}$. The subspace of all such functions whose L^p norm is finite is a normed linear space, denoted $L^p(A)$. $L^p(\mathbb{T}^q)$ will be abbreviated to L^p. The p-distance between functions $f, g \in L^p$ can now be expressed as $\|f-g\|_p$, and we will use the phrase 'convergence in the sense of L^p' when a sequence of functions (f_n) converges to f, to mean $\|f_n - f\|_p \to 0$ as $n \to \infty$. The most commonly used norms are $p = 1, 2, \infty$, representing the absolute value, Euclidean, and supremum norms respectively.

Such norms are also useful for sequence spaces, and so we will define them as done in the literature. Let $\mathbf{a} = \{a_k\}_{k=1}^{\infty}$, and $p \in [1,\infty]$. Then the ℓ^p norm of $\mathbf{a}$ is

$$|\mathbf{a}|_p := \begin{cases} \left\{\displaystyle\sum_{k=1}^{\infty} |a_k|^p\right\}^{1/p}, & \text{if } 1 \le p < \infty \quad , \\ \displaystyle\sup_{k\in\mathbb{Z}, k>0} \{|a_k|\}, & \text{if } p = \infty. \end{cases} \tag{2.1.2}$$

The subspace of all sequences whose ℓ^p norm is finite is a normed linear space, denoted ℓ^p. For a vector $\mathbf{x}$ in q dimensions, $|\mathbf{x}|_p$ is understood to evaluate the norm of the sequence which is equal to x_k for $k \le q$ and vanishes otherwise; clearly any such sequence which eventually vanishes lies in each ℓ^p.

Univariate trigonometric polynomials are the prototypical 2π periodic functions and are defined as any map of the form $x \mapsto \sum_{k\in\mathbb{Z}} a_k e^{ikx}$ where each $a_k \in \mathbb{C}$ and only finitely many a_k's are non-zero. The *degree* of a trigonometric polynomial is defined as $\max\{|k| : a_k \neq 0\}$. Notice that $\{e^{ik\circ}\}_{k\in\mathbb{Z}}$ forms an orthonormal basis for $L^2(\mathbb{T})$, under the usual inner product $\langle f, g\rangle = \int_{\mathbb{T}} f\ \overline{g}\ \mathrm{d}\mu_1^*$:

$$\int_{\mathbb{T}} e^{ikx}\overline{e^{ijx}}\mathrm{d}\mu_1^*(x) = \delta_{j,k} \quad .$$

Let $\mathbb{H}_n :=$ span $\{e^{ik\circ} : |k| \leq n\}$. Hence $\mathbb{H}_n$ contains all univariate trigonometric polynomials of degree at most n. It is a well-known fact that the trigonometric polynomials are dense in $L^p(\mathbb{T})$ for $p \in [1,\infty)$. In the supremum norm, they are dense in the subset of $L^\infty(\mathbb{T})$ consisting of (equivalence classes of) continuous functions. These are consequences of the density of continuous functions in L^p for $p \in [1,\infty)$, along with Fejér's theorem as detailed in [20]. Let us denote

$$X^p(\mathbb{T}) := \{f \in L^p(\mathbb{T}) : \mathsf{dist}\ (L^p(\mathbb{T}); f, \mathbb{H}_n) \to 0 \text{ as } n \to \infty\} \quad ,$$

the *closure* of $\bigcup_n \mathbb{H}_n$ in $L^p(\mathbb{T})$. Then evidently $X^p(\mathbb{T}) = L^p(\mathbb{T})$ when $p \in [1,\infty)$, and $X^\infty(\mathbb{T})$ is the set of continuous 2π-periodic functions on $\mathbb{T}$. This makes the trigonometric polynomials powerful tools as approximation spaces - in fact, each of these properties hold true when extended to the multivariate case.

Defining the degree of a multivariate trigonometric polynomial however, is not as immediately obvious as in the univariate case. Perhaps we may take the total degree, additive in each variable, or take the largest coordinate-wise degree. In fact if we consider the degree as lying in a q-dimensional integer lattice, then each ℓ^p

norm defines a valid way to measure it. The particular choice turns out not to be critical to the discussion, and so we will use the spherical degree, and define the set of all q-dimensional trigonometric polynomials with spherical degree not exceeding n as $\mathbb{H}_n^q := \text{span}\,\{e^{i\mathbf{k}\cdot\circ} : \mathbf{k} \in \mathbb{Z}^q, |\mathbf{k}|_2 \le n\}$. Clearly $\mathbb{H}_n = \mathbb{H}_n^1$, so the notations are consistent. Then we similarly let $X^p = L^p$ for $p \in [1, \infty)$ and denote X^∞ as the set of all continuous 2π-periodic functions on $\mathbb{T}^q$. Hence, for $p \in [1, \infty]$, X^p is the closure of the trigonometric polynomials in L^p. Herein our setting will be X^p spaces, and due to our frequent reference to the trigonometric polynomials as approximation spaces, we will use, as in [15], the following notation as shorthand for the degree of approximation (c.f. (1.1.2)):

$$E_{n,p}(f) := \mathsf{dist}\,(L^p; f, \mathbb{H}_n^q) \quad .$$

When x is positive but may not be an integer, $E_{x,p}(f)$ will mean $E_{\lfloor x \rfloor,p}(f)$.

A useful property of L^p spaces is their ordering, such that $r > p \Longrightarrow L^r \subset L^p$, and for $f \in L^r$, $\|f\|_p \le \|f\|_r$. Both of these may be shown using the Hölder inequality, which we reproduce from [26] in the case of functions on $\mathbb{T}^q$.

Proposition 2.1. *Let $p \in [1, \infty)$. Define p' such that $\frac{1}{p} + \frac{1}{p'} = 1$ when $p \neq 1$ and $p' = \infty$ otherwise. Then for $f \in L^p$, $g \in L^{p'}$, we have that $fg \in L^1$ and*

$$\|fg\|_1 \le \|f\|_p \|g\|_{p'} \quad . \tag{2.1.3}$$

The proof is found in [26] .

The analogous inclusion for the ℓ^p spaces is $r > p \Longrightarrow \ell^r \supset \ell^p$, and for $\mathbf{a} \in \ell^p$, $|\mathbf{a}|_p \ge |\mathbf{a}|_r$; notice the reversal of the ordering. As we will require it later, we additionally state the Schwarz inequality in the context of sequence spaces.

Proposition 2.2. *Let* $\mathbf{a}, \mathbf{b} \in \ell^2$*. Then* $\mathbf{ab} = \{a_k b_k\}_{k=1}^{\infty} \in \ell^1$ *and*

$$|\mathbf{ab}|_1 \leq |\mathbf{a}|_2 |\mathbf{b}|_2 \quad . \tag{2.1.4}$$

In fact a direct analogue to the Hölder inequality holds for ℓ^p spaces, and this immediately shows the result. Each of these points is shown in [26] .

The last ingredient we will require is an estimate on the norm of a certain integral called a *convolution*. This is usually defined for $f, g \in L^1$ as

$$(f * g)(\mathbf{x}) := \int_{\mathbb{T}^q} f(\mathbf{t}) g(\mathbf{x} - \mathbf{t}) \mathrm{d}\mu_q^*(\mathbf{t}) = \int_{\mathbb{T}^q} f(\mathbf{x} - \mathbf{t}) g(\mathbf{t}) \mathrm{d}\mu_q^*(\mathbf{t}) \quad .$$

Then $f * g \in L^1$ as well, and a more general result is given in this weaker form of the Young inequality for functions on $\mathbb{T}^q$, reproduced from [26] .

Proposition 2.3. *Let* $p \in [1, \infty]$*. Then if* $f \in L^p$ *and* $g \in L^1$*, we have that* $f * g \in L^p$*, with*

$$\|f * g\|_p \leq \|f\|_p \|g\|_1 \quad . \tag{2.1.5}$$

The integrable function g here is referred to as the kernel of the convolution, and many of the techniques in this theory will involve convolutions against appropriate kernels.

A note on constant conventions. In the remainder of the thesis, when $c, c_1, \ldots$ are used in formulas, they denote positive constants which may depend on any number of fixed parameters in the context in which they appear. However, they will not obfuscate any critical information, and as such do not rely for example on the target function or the dimension of the approximating space when performing an approximation. Whenever this distinction may not be clear or the constant's reliance on some parameter should be highlighted, it will be noted explicitly. The values of these constants may be different at different occurrences, even within a single formula. When we say an expression A is "on the order of" B, we mean there exist constants c, c_1 such that $cB \leq A \leq c_1 B$, and equivalently will denote this by $A \sim B$. These conventions are common in approximation theory.

2.2 Harmonic Analysis Fundamentals

The building block of harmonic analysis on $\mathbb{T}^q$ is the Fourier coefficient, traditionally defined for a function $f : \mathbb{T}^q \to \mathbb{C}$ at each integer-valued q-vector $\mathbf{k}$ as

$$\widehat{f}(\mathbf{k}) := \int_{\mathbb{T}^q} f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}\mu_q^*(\mathbf{x}) \quad .$$

Notice that

$$|\widehat{f}(\mathbf{k})| \leq \int_{\mathbb{T}^q} |f(\mathbf{x})||e^{-i\mathbf{k}\cdot\mathbf{x}}| \mathrm{d}\mu_q^*(\mathbf{x}) \leq \|f\|_1 \leq \|f\|_p \text{ for } p \in [1, \infty] \quad ,$$

and so membership in any L^p secures existence of the Fourier coefficients. Herein, we will just use membership in L^1 as a condition on a function to express that membership in any L^p is sufficient.

A (multivariate) trigonometric series is formally an expression of the form

$$\sum_{\mathbf{k}\in\mathbb{Z}^q} a_{\mathbf{k}} e^{i\mathbf{k}\cdot\circ} \quad ,$$

with complex coefficients $a_{\mathbf{k}}$. All trigonometric polynomials are trigonometric series, and further for any $f \in L^1$, the Fourier series is defined as usual as a trigonometric series with $a_{\mathbf{k}} = \widehat{f}(\mathbf{k})$. While there is no reason for this series to converge in general, we have for $f \in L^2$ that

$$f(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^q} \langle f, e^{i\mathbf{k}\cdot\circ}\rangle e^{i\mathbf{k}\cdot\mathbf{x}} = \sum_{\mathbf{k}\in\mathbb{Z}^q} \widehat{f}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \quad ,$$

due to orthonormality, with equality representing convergence in the sense of L^2. Further the Parseval relationship can be deduced:

$$\|f\|_2^2 = \int_{\mathbb{T}^q} |f(\mathbf{x})|^2 \mathrm{d}\mu_q^*(\mathbf{x}) = \int_{\mathbb{T}^q} \sum_{\mathbf{k},\boldsymbol{\ell}\in\mathbb{Z}^q} \widehat{f}(\mathbf{k})\overline{\widehat{f}(\boldsymbol{\ell})} e^{i(\mathbf{k}-\boldsymbol{\ell})\cdot\mathbf{x}} \mathrm{d}\mu_q^*(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^q} |\widehat{f}(\mathbf{k})|^2 \quad . \tag{2.2.6}$$

Now, returning to $f \in L^1$, we consider the spherical partial Fourier sum of order n, which will be denoted

$$s_n(f, \mathbf{x}) := \sum_{\mathbf{k}\in\mathbb{Z}^q, |\mathbf{k}|_2 \le n} \widehat{f}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \quad .$$

We have $s_n(f) \in \mathbb{H}_n^q$, and notice that the norm convergence $s_n(f) \to f$ is clear in L^2 for $f \in L^2$, with $\|f - s_n(f)\|_2 = E_{n,2}(f)$. Does this utility of the Fourier series extend to representing L^p functions? For the following discussion we return to the univariate setting and set $q = 1$, and then the following general results hold:

When $q = 1$, for $p \in (1, \infty)$ and $f \in L^p(\mathbb{T})$, the Fourier series of f will converge to f in the sense of $L^p(\mathbb{T})$, such that

$$\|s_n(f) - f\|_p \sim \mathsf{dist}\ (L^p(\mathbb{T}); f, \mathbb{H}_n) \text{ and so } \|s_n(f) - f\|_p \to 0 \text{ as } n \to \infty \quad . \tag{2.2.7}$$

The proof of this result is quite deep and beyond the scope of this thesis, but is treated thoroughly in [29]. In the case that $p = 1$ or $p = \infty$, the Fourier series is not guaranteed to converge at all. For the supremum norm, observe that the univariate partial Fourier sums may be written as a convolution against a familiar kernel.

$$s_n(f, x) = \sum_{k\in\mathbb{Z}, |k|\le n} \widehat{f}(k) e^{ikx} = \sum_{k\in\mathbb{Z}, |k|\le n} \int_{\mathbb{T}} f(y) e^{ik(x-y)} \mathrm{d}\mu_1^*(y) = \left(f * \sum_{k\in\mathbb{Z}, |k|\le n} e^{ik\circ} \right)(x) \quad .$$

Writing the *Dirichlet kernel* as

$$D_n(x) := \frac{1}{2} \sum_{|k| \le n-1} e^{ikx} = \frac{\sin((n-1/2)x)}{2\sin(x/2)} \quad ,$$

then yields that $s_{n-1}(f) = f * 2D_n$. It can be shown that $\|D_n\|_1 \sim \ln n$, the significance of which is that $\|D_n\|_1 \to \infty$ as $n \to \infty$. A consequence of this fact is that the operator norm of s_n is not uniformly bounded. According to the uniform boundedness principle as in [20], the class of $X^\infty(\mathbb{T})$ functions for which their Fourier series does not converge in the supremum norm is dense in $X^\infty(\mathbb{T})$ (In fact, it is a dense G_δ subset of $X^\infty(\mathbb{T})$).

Pathologies exist in the case of L^1, such as the construction by Kolmogorov of functions in L^1 for which the Fourier series diverges almost everywhere (see [29]). However, as we will show later in this chapter, the Fourier coefficients of an L^1 (or $L^1(\mathbb{T})$) function uniquely determine that function. Hence, the classical approach has been to develop summability methods which yield trigonometric polynomials determined by the Fourier coefficients of a function, such that these polynomials converge to the function.

A theme connecting these summability methods is the representation of approximants as a convolution against corresponding kernels. For instance, Fejér introduced the kernel given by $F_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} D_k(x)$. Convolution of a function $f \in X^p$ against this Fejér kernel produces the Fejér means of the Fourier series of f. As a linear operator, these means have uniformly bounded norms (see [4, 20]).

Let us now return to the multivariate setting. The spherical partial sums introduced earlier may be expressed in the form of a convolution:

$$s_n(f, \mathbf{x}) = \sum_{|\mathbf{k}|_2 \le n} \widehat{f}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} = \sum_{|\mathbf{k}|_2 \le n} \int_{\mathbb{T}^q} f(\mathbf{y}) e^{-i\mathbf{k}\cdot\mathbf{y}} e^{i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}\mu_q^*(\mathbf{y}) = \left(f * \sum_{|\mathbf{k}|_2 \le n} e^{i\mathbf{k}\cdot\circ} \right)(\mathbf{x}) \quad .$$

The sequence of kernels corresponding with the sequence of spherical partial sums also does not satisfy uniform boundedness.

Further, the Fourier coefficients alone may not express the behavior of the function - the sequence of coefficients as such may not be of use, for instance, to detect singularities. In this thesis, singularities will refer to points at which a function's (appropriately defined) smoothness changes abruptly. For instance, points at which the gradient of a function $\mathbb{R}^q \to \mathbb{C}$ is discontinuous are singularities. An aspect of wavelet analysis is developing an expansion for a function with respect to a wavelet basis, the coefficients of which may be used to detect singularities or perform other analysis dependent on local phenomena of the function. The ability to perform such analysis comes from the good *localization* of the wavelet basis elements. Localization of a function $f : \mathbb{T}^q \to \mathbb{C}$ refers to a bound of the form

$$|f(\mathbf{x})| \leq g(|\mathbf{x}|_p) \text{ for } |\mathbf{x}|_p > r \quad ,$$

for some positive real r, $p \in [1, \infty]$ and a positive decreasing function $g : [0, \infty) \to \mathbb{R}$ which vanishes at ∞. The faster g decreases to 0, the better the localization. See Figure 2.1 for an example of singularity detection, and see [3] for a discussion of wavelets and wavelet transforms.

We narrow our focus to investigate the construction and properties of sequences of linear operators which arise from convolutions against sufficiently localized kernels, are determined by the Fourier coefficients of their argument, reproduce trigonometric polynomials, and which are capable of classifying smoothness of L^p functions, appropriately defined. Historically, the work of many great analysts throughout the

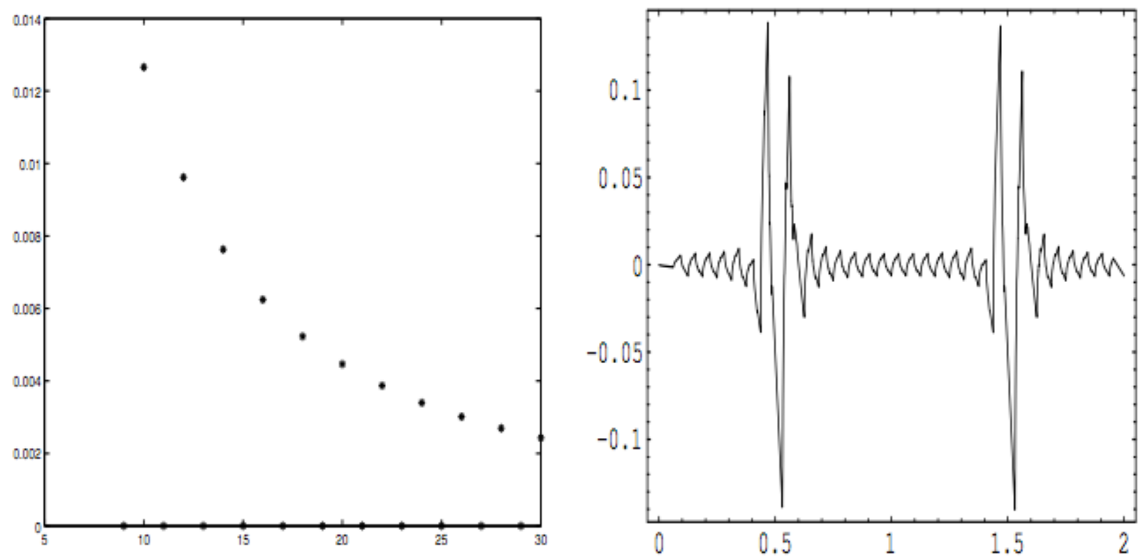

Figure 2.1: On the left is a plot of the first thirty Fourier coefficients of $f(x) = \sqrt{|\cos x|}$. On the right is the Daubechies wavelet transform $(\varphi_D^3)_4 * f$, whose peaks detect the cusps at $\pi/2$ and $3\pi/2$. Note the domain in this example is $[0, 2\pi]$, and the plot's x-axis counts multiples of π. Plots found in [22].

19^{th} and 20^{th} centuries contributed to the introduction of sophisticated summability methods, and elucidated which properties of the arising summability operators were desirable. For instance, the *shifted average* operators due to de la Vallée-Poussin have uniformly bounded operator norms independent of order, and reproduce trigonometric polynomials [21]. Mhaskar and Prestin have studied many such summability kernels, and analyzed their localization properties, which we are interested in. Some relevant papers are [21][22][23].

The ideas of Mhaskar and his collaborators which we focus on, detailed in [15] , involve constructing linear operators which, in addition to the points mentioned, are sufficiently localized to classify the *local* smoothness of L^p functions, appropriately defined. Further, they inform the construction of related *band-pass* operators which form a frame expansion, which converges to f in L^p. These criteria form the core of the wavelet-like analysis; other common names for this subject are Littlewood-Paley theory (after two of the founding analysts in the area) and multiscale analysis. The term "wavelet-like" refers to the fact mentioned earlier, that wavelet coefficients may be used to classify local smoothness of a function in conditions where the Fourier coefficients alone cannot. Daubechies discusses this property of wavelets in Chapter 9 of [3], and we seek a characterization in the spirit of Jaffard's result [3, Theorem 9.2.1]. Hence, as the operators discussed classify local smoothness as well, and are defined in terms of Fourier coefficients, we call the expansion wavelet-like. This is detailed in the next sections.

2.3 Low-Pass Filters and the Summability Operator

We wish to construct a sufficiently general class of localized kernels, and will define more precisely what is meant by smoothness and by localization. To this effort, we follow the example of using values of a compactly-supported function $H : \mathbb{R}^q \to \mathbb{R}$ in our construction of a kernel

$$\mathbf{x} \mapsto \sum_{\mathbf{k} \in \mathbb{Z}^q} H(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \quad ,$$

as written in [15] and pioneered in the work of Hardy. Recall that the support of a real or complex-valued function f, denoted by $\mathrm{supp}(f)$, is the closure of the set $\{\mathbf{x} : f(\mathbf{x}) \neq 0\}$, and we say f "is supported on" A to mean $\mathrm{supp}(f) \subset A$. Hence H being compactly supported makes the seemingly infinite sum above finite

- a criterion equivalent to compact support is the existence of some b such that $|\mathbf{x}|_2 > b \Longrightarrow H(\mathbf{x}) = 0$. Next we introduce the necessary definitions and notations for differentiation in the multivariate setting.

For a positive integer S, we say a function $H : \mathbb{R}^q \to \mathbb{C}$ is S-times continuously differentiable when for $\mathbf{r} \in \mathbb{Z}_+^q$, $|\mathbf{r}|_1 \leq S$, $D^{\mathbf{r}}H$ exists and is continuous. Then we define the expression

$$\|H\|_{D^S} = \max_{\mathbf{r}\in\mathbb{Z}_+^q, |\mathbf{r}|_1 = S} \|D^{\mathbf{r}}H\|_1 \quad ,$$

even when the expression is not finite.

We now focus our attention on a certain class of compactly supported, smooth functions which will facilitate the development of kernels with sufficiently high localization.

Definition 2.4. *Let S be a positive integer. Then $h : \mathbb{R} \to \mathbb{R}$ is called an S-smooth* ***low-pass filter*** *when the following are satisfied:*

- *h is S-times continuously differentiable.*
- *h is even and bounded.*
- *h is non-increasing on $[0, \infty)$.*
- *$h(x) = 1$ when $|x| \leq 1/2$ and $h(x) = 0$ when $|x| \geq 1$.*

Example 2.3.1. For any non-negative integer S we may generate an S-smooth low-pass filter h_S by writing the expression

$$h_S(x) := A_S \int_x^1 (2t-1)^S (1-t)^S \mathrm{d}t \quad \text{for } x \in \left(\frac{1}{2}, 1\right) \quad ,$$

where A_S is a normalizing constant,

$$A_S = 2^{S+1}\frac{(2S+1)!}{(S!)^2} \quad .$$

To satisfy the conditions for being a low-pass filter, additionally $h_S(x) = 1$ on $\left[0, \frac{1}{2}\right]$, $h_S(x) = 0$ on $[1, \infty)$, and $h_S(-x) = h_S(x)$. $\square$

EXAMPLE 2.3.2. The function h_∞ which satisfies the conditions for being a low-pass filter and is defined on $\left(\frac{1}{2}, 1\right)$ by

$$h_\infty(x) = 2\left[1 + \exp\left(\frac{(2x-1)^2}{1-(2x-1)^2}\right)\right]^{-1} \qquad \text{for } x \in \left(\frac{1}{2}, 1\right) \quad .$$

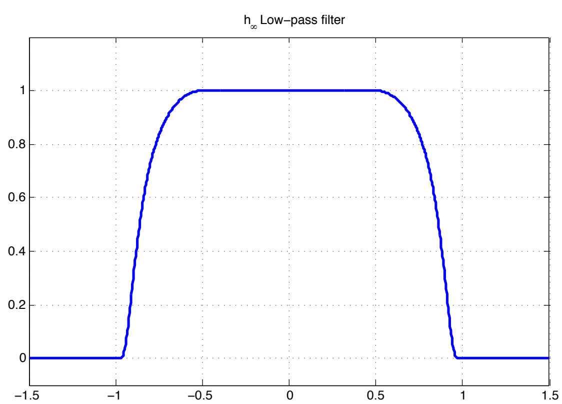

Figure 2.2: This low-pass filter is S-smooth for every positive integer S.

$\square$

We may construct a series of kernels with localization as nice as we would like by considering, for each positive integer n, a choice of H given by $\mathbf{x} \mapsto h(|\mathbf{x}|_2/n)$ for an S-smooth low-pass filter h. Clearly each such function will be compactly supported, and as we will see, the localization is controlled by S.

Definition 2.5. *Let S be a positive integer, and h an S-smooth low-pass filter. Then* *we may define the low-pass* ***summability kernel*** *by*

$$\Phi_n(h, \mathbf{t}) := \sum_{\mathbf{k}\in\mathbb{Z}^q} h\left(\frac{|\mathbf{k}|_2}{n}\right) e^{i\mathbf{k}\cdot\mathbf{t}} \quad .$$

The sum expression $\sum_{\mathbf{k}\in\mathbb{Z}^q}$ is really $\sum_{|\mathbf{k}|_2\leq n}$, and so the sum is finite. The critical property of this kernel is the localization estimate it satisfies, as follows.

Theorem 2.6. *Let $S > q$ be a positive integer, and h an S-smooth low-pass filter. Then*

$$|\Phi_n(h, \mathbf{x})| \leq cn^q / \max\left\{1, (n|\mathbf{x}|_1)^S\right\} \text{ for all } \mathbf{x} \in \mathbb{T}^q \quad , \tag{2.3.8}$$

and

$$\|\Phi_n(h)\|_1 \leq c \quad , \tag{2.3.9}$$

where the constant c will depend on h but not on n.

A consequence of the estimate (2.3.8) is that for any $\gamma > 0$, when $S > q+\gamma$ and h is an S-smooth low-pass filter, $\Phi_n(h, \mathbf{x})$ eventually vanishes faster than $(n|\mathbf{x}|_1)^{-\gamma}$. This is an important aspect of having good localization, as demonstrated in Figure 2.3. Note that h_0 as in Example 2.3.1 yields the classical *de la Vallée-Poussin kernel* $\Phi_n(h_0)$.

The proof of this estimate relies on the multivariate Poisson summation formula, which itself relies on the properties of the Fourier transform. Hence we will introduce these concepts and the accompanying notation here. Much of this exposition is taken directly from the corresponding discussion in [15].

For $F \in L^1(\mathbb{R}^q)$, the Fourier transform is defined by

$$\widehat{F}(\mathbf{x}) = \frac{1}{(2\pi)^q} \int_{\mathbb{R}^q} F(\mathbf{y}) \exp(-i\mathbf{y}\cdot\mathbf{x}) d\mathbf{y}, \qquad \mathbf{x} \in \mathbb{R}^q \quad ,$$

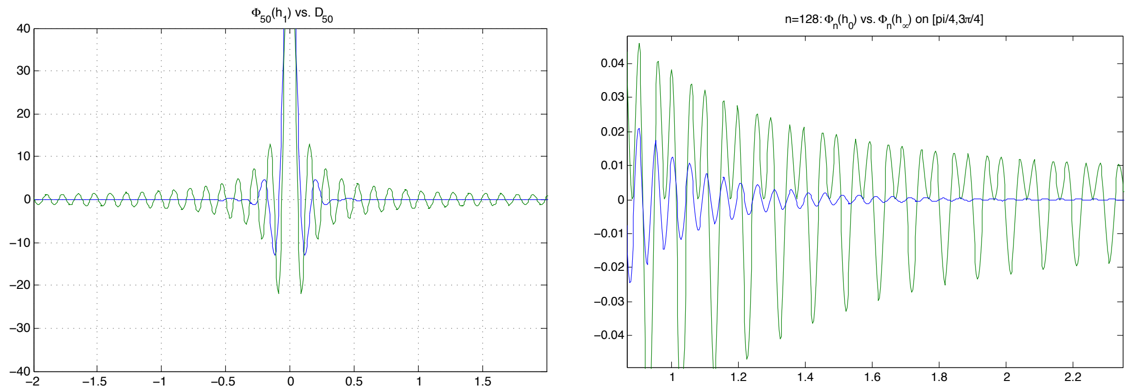

Figure 2.3: Some plots of localized kernels in the univariate case. On the left, notice how the summability kernel for a 1-smooth low-pass filter has oscillations which vanish extremely quickly, unlike the Dirichlet kernel (in green). On the right, the summability kernel for a 0-smooth low-pass filter is plotted in green against that of an infinitely smooth filter in blue, and the graph is shown on $[\pi/4, 3\pi/4]$.

and the inverse Fourier transform by

$$\widetilde{F}(\mathbf{x}) = \int_{\mathbb{R}^q} F(\mathbf{y}) \exp(i\mathbf{y} \cdot \mathbf{x}) \mathrm{d}\mathbf{y}, \qquad \mathbf{x} \in \mathbb{R}^q \quad .$$

As we denoted the Fourier coefficients of an L^1 function f by $\widehat{f}(\mathbf{m})$, there is an abuse of notation here. $\widehat{F}(\mathbf{m})$ denotes the Fourier transform of the $L^1(\mathbb{R}^q)$ function F evaluated at $\mathbf{m}$. This is however a very common abuse, and the meaning should be clear from context.

Proposition 2.7. (a) *If both* F *and* $\widehat{F}$ *are in* $L^1(\mathbb{R}^q)$*, then the* Fourier inversion formula *holds for every* $\mathbf{x} \in \mathbb{R}^q$*:*

$$F(\mathbf{x}) = \widetilde{\widehat{F}}(\mathbf{x}) = \widehat{\widetilde{F}}(\mathbf{x}) \quad .$$

(b) *Let $F \in L^1(\mathbb{R}^q)$. Then we have*

$$\|F\|_{1,\mathbb{R}^q} = \sum_{\mathbf{k}\in\mathbb{Z}^q} \|F(\circ + 2\pi\mathbf{k})\|_1 < \infty.$$

Particularly, the function

$$f(\mathbf{x}) := \sum_{\mathbf{k}\in\mathbb{Z}^q} F(\mathbf{x} + 2\pi\mathbf{k}) \tag{2.3.10}$$

is defined for almost all $\mathbf{x} \in \mathbb{T}^q$, $f \in L^1$, and $\|f\|_1 = \|F\|_{1,\mathbb{R}^q}$. Moreover, $\widehat{f}(\mathbf{k}) = \widehat{F}(\mathbf{k})$. If both the series defining f and the Fourier series of f converges uniformly on $\mathbb{T}^q$, then we have the following Poisson summation formula *valid for $\mathbf{x} \in \mathbb{T}^q$:*

$$\sum_{\mathbf{m}\in\mathbb{Z}^q} \widehat{F}(\mathbf{m}) \exp(i\mathbf{m}\cdot\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^q} F(\mathbf{x} + 2\pi\mathbf{k}) \quad . \tag{2.3.11}$$

(c) *If $S \geq 1$ is an integer, F is compactly supported, and is S times continuously differentiable then for $\mathbf{x} \in \mathbb{R}^q - \mathbf{0}$,*

$$|\widehat{F}(\mathbf{x})| \leq \frac{q^S \|F\|_{D^S}}{|\mathbf{x}|_1^S} \quad \textit{and} \quad |\widetilde{F}(\mathbf{x})| \leq \frac{(2\pi)^q q^S \|F\|_{D^S}}{|\mathbf{x}|_1^S} \quad . \tag{2.3.12}$$

Proof. Part (a) is given in [28, Chapter 1, Corollary 1.21]; the proof involves details of measure theory which will run this thesis off-topic and will not be reproduced. For part (b), each term $\|F(\circ + 2\pi\mathbf{k})\|_1$ in the series is just $\int_{A_\mathbf{k}} |F(\mathbf{x})| \mathrm{d}\mathbf{x}$ where $A_\mathbf{k} = \Pi_{j=1}^q [-\pi + 2\pi k_j, \pi + 2\pi k_j]$. The $A_\mathbf{k}$'s partition $\mathbb{R}^q$, and so $\sum_{\mathbf{k}\in\mathbb{Z}^q} \|F(\circ + 2\pi\mathbf{k})\|_1 = \|F\|_{1,\mathbb{R}^q} < \infty$ by hypothesis. Hence $f \in L^1$ with $\|f\|_1 = \|F\|_{1,\mathbb{R}^q}$ as desired. To obtain the Poisson summation formula, use the Dominated Convergence Theorem (for the Lebesgue measure) to obtain that $\widehat{f}(\mathbf{m}) = \widehat{F}(\mathbf{m})$ for $\mathbf{m} \in \mathbb{Z}^q$. Then of course when both series converge uniformly the equality (2.3.11) holds. For part (c), let F be compactly supported and S-times continuously differentiable. Observe that if we integrate the expression for the Fourier transform by parts with respect to x_j, we obtain

$$(-ix_j)\widehat{F}(\mathbf{x}) = \frac{1}{(2\pi)^q}\int_{\mathbb{R}^q}(\partial_j F)(\mathbf{y})\exp(-i\mathbf{y}\cdot\mathbf{x})\mathrm{d}\mathbf{y} \quad .$$

A repeated application of this formula shows that for $\mathbf{r}\in\mathbb{Z}_+^q$, $|\mathbf{r}|_1 \le S$,

$$(-i\mathbf{x})^{\mathbf{r}}\widehat{F}(\mathbf{x}) = \frac{1}{(2\pi)^q}\int_{\mathbb{R}^q}(D^{\mathbf{r}}F)(\mathbf{y})\exp(-i\mathbf{y}\cdot\mathbf{x})d\mathbf{y} \quad .$$

The notation $\mathbf{v}^{\mathbf{r}}$ refers to the vector whose components are $v_j^{r_j}$. Additionally, when $|\mathbf{r}|_1 = S$, let $\binom{S}{\mathbf{r}} := S!/r_1!r_2!\dots r_q!$. It follows that for every $\mathbf{x}\in\mathbb{R}^q$,

$$|\mathbf{x}|_1^S|\widehat{F}(\mathbf{x})| = \sum_{\mathbf{r}\in\mathbb{Z}_+^q,\ |\mathbf{r}|_1=S}\binom{S}{\mathbf{r}}|\mathbf{x}|^{\mathbf{r}}|\widehat{F}(\mathbf{x})| \le \sum_{\mathbf{r}\in\mathbb{Z}_+^q,\ |\mathbf{r}|_1=S}\binom{S}{\mathbf{r}}\|D^{\mathbf{r}}F\|_{1,\mathbb{R}^q} \le q^S\|F\|_S \quad .$$

The second estimate in (c) follows from the fact that $\widetilde{F}(\mathbf{x}) = (2\pi)^q\widehat{F}(-\mathbf{x})$. $\square$

We will need to solve integrals of the form $\int_{\mathbb{R}^q} f(|\mathbf{x}|_1)\mathrm{d}\mathbf{x}$, and so we reproduce Lemma 2.2.1 from [15].

Lemma 2.8. *For any $f:[0,\infty)\to\mathbb{R}$ for which the mapping $t\mapsto f(t)t^{q-1}$ is integrable on $[0,\infty)$, we have*

$$\int_{\mathbb{R}^q} f(|\mathbf{x}|_1)d\mathbf{x} = \frac{2^q}{(q-1)!}\int_0^\infty f(t)t^{q-1}dt. \qquad (2.3.13)$$

Proof. We must make a change of variables - to keep the Jacobian determinant simple, use

$$x_j = tt_j, \quad j=1,\cdots,q-1, \quad x_q = t(1-t_1-\cdots-t_{q-1}) \quad .$$

Then, changing variables, we deduce that

$$\int_{\mathbb{R}^q} f(|\mathbf{x}|_1)\mathrm{d}\mathbf{x} = 2^q\int_{[0,\infty)^q} f(|\mathbf{x}|_1)\mathrm{d}\mathbf{x} = 2^q\int_{\{\mathbf{t}\in\mathbb{R}_+^{q-1},\ |\mathbf{t}|_\infty\le 1,\ |\mathbf{t}|_1=1\}}\int_0^\infty f(t)t^{q-1}dtd\mathbf{t}.$$

It is easily verified by induction that

$$\int_{\{\mathbf{t}\in\mathbb{R}_+^{q-1},\ |\mathbf{t}|_\infty\le 1,\ |\mathbf{t}|_1=1\}} d\mathbf{t} = \frac{1}{(q-1)!} \quad .$$

The result (2.3.13) follows. $\square$

As a consequence of this lemma, we have for every $r > 0$,

$$\int_{|\mathbf{x}|_1\le r} \mathrm{d}\mathbf{x} = \frac{2^q}{(q-1)!}\int_0^r t^{q-1}dt = \frac{2^q r^q}{q!} \quad , \tag{2.3.14}$$

and when $S > q$,

$$\int_{|\mathbf{x}|_1\ge r} \frac{\mathrm{d}\mathbf{x}}{|\mathbf{x}|_1^S} = \frac{2^q}{(q-1)!}\int_r^\infty t^{q-1-S}dt = \frac{2^q r^q}{q!}\frac{q}{(S-q)r^S} \quad . \tag{2.3.15}$$

Now we will prove Theorem 2.6.

Proof. Let $H_n(\mathbf{x}) = h\left(\frac{|\mathbf{x}|_2}{n}\right)$. As H_n is continuous and compactly supported, $H_n \in L^1(\mathbb{R}^q)$. $\Phi_n(h) \in \mathbb{H}_n^q$, and so its series representation is finite, and in particular absolutely and uniformly convergent. Further, we have $H_n(\mathbf{x}) = H_1(\mathbf{x}/n)$. Notice that since h is constant on a neighborhood of $\mathbf{0}$,

$$\nabla H_1(\mathbf{x}) = \frac{\mathbf{x}}{|\mathbf{x}|_2}h'(|\mathbf{x}|_2) = 0 \text{ on a nbhd. of } \mathbf{0}.$$

Hence, H_1 (and so H_n) inherits being $S > q$ times continuously differentiable from the fact that h is S-smooth. So, we use (2.3.12) and Lemma 2.8 to deduce that $\widehat{H}_n \in L^1(\mathbb{R}^q)$. Hence, both the Fourier inversion formula and the Poisson summation formula hold for all $\mathbf{x} \in \mathbb{R}^q$, and we have for $\mathbf{x} \in \mathbb{R}^q$,

$$\Phi_n(h,\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^q} \widetilde{H}_n(\mathbf{x}+2\mathbf{k}\pi) \quad .$$

Notice that

$$\widetilde{H}_n(\mathbf{x}) = \int_{\mathbb{R}^q} H_n(\mathbf{y}) e^{i\mathbf{x}\cdot\mathbf{y}} \mathrm{d}\mathbf{y} = \int_{\mathbb{R}^q} H_1(\mathbf{y}/n) e^{i\mathbf{x}\cdot\mathbf{y}} \mathrm{d}\mathbf{y} = n^q \int_{\mathbb{R}^q} H_1(\mathbf{y}) e^{in\mathbf{x}\cdot\mathbf{y}} \mathrm{d}\mathbf{y} = n^q \widetilde{H}_1(n\mathbf{x}) \ ,$$

and so

$$\Phi_n(h, \mathbf{x}) = n^q \sum_{\mathbf{k}\in\mathbb{Z}^q} \widetilde{H}_1\left(n(\mathbf{x} + 2\mathbf{k}\pi)\right) \quad .$$

Enlisting the fact that H_1 is S times continuously differentiable, we now use (2.3.12) to obtain for $\mathbf{x} \in \mathbb{T}^q \setminus \{\mathbf{0}\}$,

$$\begin{aligned} |\Phi_n(h, \mathbf{x})| &\leq (2\pi)^q q^S n^q \|H_1\|_{D^S} \sum_{\mathbf{k}\in\mathbb{Z}^q} \frac{n^{-S}}{|\mathbf{x} + 2\mathbf{k}\pi|_1^S} \\ &= c(q, h) n^{q-S} \left\{ \frac{1}{|\mathbf{x}|_1^S} + \sum_{\mathbf{k}\in\mathbb{Z}^q,\ |\mathbf{k}|_\infty \geq 1} \frac{1}{|\mathbf{x} + 2\mathbf{k}\pi|_1^S} \right\} \quad . \end{aligned}$$

We will not worry about the exact value of the constant except to mention that it is finite, and to recall its dependence on the dimension and choice of filter. Since $|\mathbf{x}|_\infty \leq \pi$, we obtain for $\mathbf{k} \in \mathbb{Z}^q$, $|\mathbf{k}|_\infty \geq 1$, that

$$|\mathbf{x} + 2\mathbf{k}\pi|_1 \geq |\mathbf{x} + 2\mathbf{k}\pi|_\infty \geq 2\pi|\mathbf{k}|_\infty - |\mathbf{x}|_\infty \geq |\mathbf{x}|_\infty(2|\mathbf{k}|_\infty - 1) \geq \frac{|\mathbf{x}|_1}{q}(2|\mathbf{k}|_\infty - 1) \quad .$$

Therefore, for $\mathbf{x} \in \mathbb{T}^q \setminus \{\mathbf{0}\}$,

$$\begin{aligned} \sum_{\substack{\mathbf{k}\in\mathbb{Z}^q \\ |\mathbf{k}|_\infty \geq 1}} |\mathbf{x} + 2\mathbf{k}\pi|_1^{-S} &\leq \frac{q^S}{|\mathbf{x}|_1^S} \sum_{j=1}^{\infty} \sum_{|\mathbf{k}|_\infty = j} (2|\mathbf{k}|_\infty - 1)^{-S} \\ &= \frac{q^S}{|\mathbf{x}|_1^S} \sum_{j=1}^{\infty} |\{\mathbf{k} \in \mathbb{Z}^q : |\mathbf{k}|_\infty = j\}| (2j - 1)^{-S} \quad . \quad (2.3.16) \end{aligned}$$

For $\mathbf{k} \in \mathbb{Z}^q$, $j \geq 1$, $|\mathbf{k}|_\infty = j$ if and only if for some $\ell \in \{1, \cdots, q\}$, there is a subset $U \subseteq \{1, \cdots, q\}$ such that $|U| = \ell$, $k_m = \pm j$ for $m \in U$, and $|k_m| \leq j - 1$ for $m \in \{1, \cdots, q\} - U$. We then see that

$$|\{\mathbf{k} \in \mathbb{Z}^q : |\mathbf{k}|_\infty = j\}| = \sum_{\ell=1}^{q} \binom{q}{\ell} 2^\ell (2j - 1)^{q-\ell} = \sum_{\ell=0}^{q-1} \binom{q}{\ell} 2^{q-\ell} (2j - 1)^\ell.$$

Hence, since $S > q$, the infinite series in (2.3.16) converges. Thus we have for $\mathbf{x} \in \mathbb{T}^q \setminus \{\mathbf{0}\}$,

$$|\Phi_n(h, \mathbf{x})| \leq c(q, h) n^q \left(n|\mathbf{x}|_1\right)^{-S} \quad .$$

When $\mathbf{x} = \mathbf{0}$,

$$|\Phi_n(h, \mathbf{0})| \leq \sum_{\mathbf{k} \in \mathbb{Z}^q} |H_n(\mathbf{k})| = \sum_{\mathbf{k} \in \mathbb{Z}^q} \left| h\left(\frac{|\mathbf{k}|_2}{n}\right)\right| \leq cn^q \quad ,$$

the number of integer grid points in the q-sphere with radius n. Finally, we may conclude that for $\mathbf{x} \in \mathbb{T}^q$, (2.3.8) holds.

In light of this, to bound $\|\Phi_n(h)\|_1$, we split the integral along the boundary $|\mathbf{x}|_1 = \frac{1}{n}$:

$$\begin{aligned} \|\Phi_n(h)\|_1 &\leq \int_{|\mathbf{x}|_1 \leq 1/n} |\Phi_n(h, \mathbf{x})| \mathrm{d}\mu_q^*(\mathbf{x}) + \int_{|\mathbf{x}|_1 \geq 1/n} |\Phi_n(h, \mathbf{x})| \mathrm{d}\mu_q^*(\mathbf{x}) \\ &\leq cn^q \int_{|\mathbf{x}|_1 \leq 1/n} \mathrm{d}\mathbf{x} + c_0 n^{q-S} \int_{|\mathbf{x}|_1 \geq 1/n} \frac{\mathrm{d}\mathbf{x}}{|\mathbf{x}|_1^S} \quad , \end{aligned}$$

noting that we change from integration over $\mathbb{T}^q$ in the first line to integration over $\mathbb{R}^q$ in the second. Recalling that $S > q$, and using (2.3.14) and (2.3.15), we conclude that

$$\int_{|\mathbf{x}|_1 \leq 1/n} \mathrm{d}\mathbf{x} = c_1(q) n^{-q} \text{ and } \int_{|\mathbf{x}|_1 \geq 1/n} \frac{\mathrm{d}\mathbf{x}}{|\mathbf{x}|_1^S} = c_2(q, h) n^{S-q} \quad .$$

The exact constants are not vital here. However, this is enough to give that $\|\Phi_n(h)\|_1$ is bounded by a constant, as desired. □

Convolving against this kernel produces the linear operator which is central to this chapter.

Definition 2.9. *Let S be a positive integer, and h an S-smooth low-pass filter. Let $p \in [1, \infty]$, and $f \in X^p$. Then the low-pass* ***summability operator*** *is given by*

$$\sigma_n^*(h, f, \mathbf{x}) := (f * \Phi_n(h))(\mathbf{x}) = \int_{\mathbb{T}^q} \Phi_n(h, \mathbf{x} - \mathbf{t}) f(\mathbf{t}) \mathrm{d}\mu_q^*(\mathbf{t}) \quad .$$

Linearity follows from linearity of the integral. A quick calculation then gives, by using the definition of Φ_n and of the Fourier coefficients,

$$\sigma_n^*(h, f, \mathbf{x}) = \sum_{|\mathbf{k}|_2 \le n} h\left(\frac{|\mathbf{k}|_2}{n}\right) e^{i\mathbf{k}\cdot\mathbf{x}} \int_{\mathbb{T}^q} f(\mathbf{t}) e^{-i\mathbf{k}\cdot\mathbf{t}} \mathrm{d}\mu_q^*(\mathbf{t}) = \sum_{|\mathbf{k}|_2 \le n} h\left(\frac{|\mathbf{k}|_2}{n}\right) \widehat{f}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \quad .$$

Hence the term low-pass filter refers to the fact that the action of h in this convolution is to reproduce low frequencies while suppressing high frequencies – we incorporate smoothness and being nonincreasing into the definition in order to streamline later results, and because we do not seek full generality here.

An immediate consequence of this definition is that σ_{2n}^* reproduces trigonometric polynomials of degree n. If $T \in \mathbb{H}_n^q$, we have, by orthogonality or by observation, that the Fourier coefficients of T are simply its trigonometric polynomial coefficients; that is, $T(\mathbf{x}) = \sum_{|\mathbf{k}|_2 \le n} \widehat{T}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}}$, and $|\mathbf{k}|_2 > n \Longrightarrow \widehat{T}(\mathbf{k}) = 0$. Hence,

$$\begin{aligned} \sigma_{2n}^*(h, T, \mathbf{x}) &= \sum_{|\mathbf{k}|_2 \le 2n} h\left(\frac{|\mathbf{k}|_2}{2n}\right) \widehat{T}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \\ &= \sum_{|\mathbf{k}|_2 \le n} h\left(\frac{|\mathbf{k}|_2}{2n}\right) \widehat{T}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} = \sum_{|\mathbf{k}|_2 \le n} \widehat{T}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} = T(\mathbf{x}) \quad . \end{aligned}$$

Notice that, again, the choice of filter here played no part, so long as it satisfied the definition – we merely need that $h(x) = 1$ for $x \le 1/2$. While specific constants in formulas will in general rely on the choice of filter, the restrictions on the filter in the theory are generally only on its S-smoothness.

The operator norm of the summability operators is uniformly bounded, as desired.

Proposition 2.10. *Let $S > q$ be a positive integer, and h an S-smooth low-pass filter. Let $p \in [1, \infty]$, and $f \in X^p$. Then*

$$\|\sigma_n^*(h, f)\|_p \leq c\|f\|_p \quad . \tag{2.3.17}$$

Proof. Via the Young inequality (2.1.5) and the uniform boundedness of Φ_n (2.3.9),

$$\|\sigma_n^*(h, f)\|_p = \|\Phi_n(h) * f\|_p \leq \|\Phi_n(h)\|_1 \|f\|_p \leq c\|f\|_p \quad .$$

□

These two observations, of uniform boundedness and polynomial reproduction, allow us to fulfill the promise of the second section, that the Fourier coefficients of an L^p function determine it uniquely. This is because the summability operators converge to the X^p function they operate on in the sense of L^p.

Proposition 2.11. *Let $S > q$ be a positive integer, and h an S-smooth low-pass filter. Let $p \in [1, \infty]$, and $f \in X^p$. Then*

$$E_{2n,p}(f) \leq \|\sigma_{2n}^*(h, f) - f\|_p \leq cE_{n,p}(f) \quad , \tag{2.3.18}$$

and in particular $\sigma_n^*(h, f) \to f$ as $n \to \infty$.

Proof. The first inequality is immediate from the fact that $\sigma_{2n}^*(h, f) \in \mathbb{H}_{2n}^q$. Now, let $T \in \mathbb{H}_n^q$. Then we have

$$\|\sigma_{2n}^*(h, f) - f\|_p = \|\sigma_{2n}^*(h, f) - T + T - f\|_p \leq \|\sigma_{2n}^*(h, f) - T\|_p + \|T - f\|_p \quad .$$

Next we employ the fact that σ_{2n} reproduces $T \in \mathbb{H}_n^q$, along with linearity and (2.3.17), to deduce that

$$\|\sigma_{2n}^*(h, f) - T\|_p = \|\sigma_{2n}^*(h, f - T)\|_p \leq c\|f - T\|_p \quad .$$

Thus, $\|\sigma_{2n}^*(h,f) - f\|_p \le (c+1)\|f - T\|_p$. But T was arbitrary, so for any $\epsilon > 0$, we may choose T_ϵ with $\|f - T_\epsilon\|_p < E_{n,p}(f) + \epsilon$. Hence $\|\sigma_{2n}^*(h,f) - f\|_p \le cE_{n,p}(f)$ as desired.

L^p convergence follows from the definition of X^p, specifically the density of trigonometric polynomials in X^p. $\square$

Let $\widehat{f_1}(\mathbf{k}) = \widehat{f_2}(\mathbf{k})$ for each $\mathbf{k} \in \mathbb{Z}^q$. As the summability operators use only the Fourier coefficients in their definition, given an integer $S > q$, for a fixed S-smooth low-pass filter h it follows that $\sigma_n^*(h, f_1) = \sigma_n^*(h, f_2)$ for each n. Suppose $f_1 \in X^p$ for some $p \in [1, \infty]$. Then $\sigma_n^*(h, f_1) \to f_1$ as $n \to \infty$, with convergence in the sense of L^p. Just the same, $\sigma_n^*(h, f_2) \to f_1$ as $n \to \infty$, and $\sigma_n^*(h, f_2) \to f_2$ as $n \to \infty$. Taken together this means that $f_1 = f_2$ a.e., and so represent the same equivalence class.

The rest of the results we explore in this chapter will primarily be quoted from [15] or specific instances of theorems therein, and underscore the success of wavelet-like representation for periodic functions in X^p, expressed entirely in terms of Fourier coefficients.

2.4 Multvariate Approximation Results

In the case of univariate functions, the well-known Favard estimate provides a relationship between the smoothness of a function in $X^p(\mathbb{T})$ in terms of number of derivatives, and the degree of approximation achievable by trigonometric polynomials. The Bernstein inequality may be used to estimate the norm of a trigonometric polynomial's derivatives, and together they provide the backbone for deriving the direct and converse theorems of approximation by trigonometric polynomials on $\mathbb{T}$. We state them here without proof, referencing Theorem 1.1.1 of [15].

Theorem 2.12. *Let $1 \leq p \leq \infty$, $r, n \geq 1$ be integers. $\| \circ \|_p$ denotes the $L^p(\mathbb{T})$ norm in this theorem only.*

(a) (Favard estimate) *If $f \in X^p(\mathbb{T})$ has $r-1$ absolutely continuous derivatives, and $f^{(r)} \in L^p(\mathbb{T})$, then*

$$\mathit{dist}\ (L^p(\mathbb{T}); f, \mathbb{H}_n) \leq \frac{c}{(n+1)^r} \mathit{dist}\ (L^p(\mathbb{T}); f^{(r)}, \mathbb{H}_n) \leq \frac{c}{(n+1)^r} \|f^{(r)}\|_p \quad . \tag{2.4.19}$$

(b) (Bernstein inequality) *If $T \in \mathbb{H}_n$ then*

$$\|T^{(r)}\|_p \leq n^r \|T\|_p. \tag{2.4.20}$$

A multivariate analogue exists for each, using the appropriate notion of differentiability. For the remainder of the chapter, fix an S-smooth low-pass filter h. Conditions on S will be specified in each instance where h is required. Note that from the construction in Example 2.3.1, for any positive integer S we may find an S-smooth low-pass filter. Constants introduced may depend on specific choice of h.

First we mention a necessary property of the Fourier coefficients of a convolution.

Proposition 2.13. *The Fourier coefficients of a convolution of two functions $f, g \in L^1$ are given for $\mathbf{k} \in \mathbb{Z}^q$ as*

$$\widehat{(f * g)}(\mathbf{k}) = \widehat{f}(\mathbf{k})\widehat{g}(\mathbf{k}) \quad .$$

Proof. This is a consequence of Fubini's theorem. Applying the appropriate definitions and a change of variables directly yields

$$\begin{aligned}
\widehat{(f*g)}(\mathbf{k}) &= \int_{\mathbb{T}^q} (f*g)(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}\mu_q^*(\mathbf{x}) \\
&= \int_{\mathbb{T}^q}\int_{\mathbb{T}^q} f(\mathbf{y}) g(\mathbf{x}-\mathbf{y}) e^{-i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}\mu_q^*(\mathbf{y}) \mathrm{d}\mu_q^*(\mathbf{x}) \\
&= \int_{\mathbb{T}^q} f(\mathbf{y}) e^{-i\mathbf{k}\cdot\mathbf{y}} \int_{\mathbb{T}^q} g(\mathbf{x}-\mathbf{y}) e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \mathrm{d}\mu_q^*(\mathbf{y}) \mathrm{d}\mu_q^*(\mathbf{x}) \\
&= \int_{\mathbb{T}^q} f(\mathbf{y}) e^{-i\mathbf{k}\cdot\mathbf{y}} \mathrm{d}\mu_q^*(\mathbf{y}) \int_{\mathbb{T}^q} g(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} \mathrm{d}\mu_q^*(\mathbf{x}) = \widehat{f}(\mathbf{k})\widehat{g}(\mathbf{k}) \text{ as desired.}
\end{aligned}$$

□

We will denote the Laplacian operator by $\Delta := \sum_{j=1}^{q} \partial_j^2$, and observe that for a sufficiently smooth f,

$$\widehat{(-\Delta f)}(\mathbf{k}) = |\mathbf{k}|_2^2 \widehat{f}(\mathbf{k}) \quad ,$$

for each $k \in \mathbb{Z}^q$. For positive even integer r, we will then define the differential operator $(-\Delta)^{r/2}$ formally by

$$\widehat{(-\Delta)^{r/2}} f(\mathbf{k}) := |\mathbf{k}|_2^r \widehat{f}(\mathbf{k}) \tag{2.4.21}$$

for $f \in L^1$ and $\mathbf{k} \in \mathbb{Z}^q$. We will denote W_r^p the space of all functions $f \in X^p$ such that $(-\Delta)^{r/2} f \in X^p$. Note that all trigonometric polynomials of any degree lie in each W_r^p trivially, since any trigonometric polynomial is its own Fourier series.

With respect to these operators, the analogous Favard estimate and Bernstein inequality each take a remarkably simple form, expressing clearly the dependence of the degree of approximation by polynomials on the smoothness of the target function, in terms of derivatives.

Theorem 2.14. .

(a) Let $p \in [1, \infty]$, $f \in X^p$, and r be a positive even integer. Then if additionally

$f \in W_r^p$,

$$E_{n,p}(f) \leq cn^{-r}\|(-\Delta)^{r/2}f\|_p \quad .$$

(b) Let $T \in \mathbb{H}_n^q$ and r be a positive even integer. Then for each $p \in [1, \infty]$,

$$\|(-\Delta)^{r/2}T\|_p \leq cn^r\|T\|_p \quad .$$

The proofs of both estimates can be achieved using the properties of the low-pass summability kernels and operators introduced in the last section, due to their high localization. [15, Chapter 2, Theorem 3.3] gives a general treatment, from which the results follow.

The univariate case provides a sharp example of when the number of derivatives alone is a strong enough smoothness criterion to classify the degree of approximation: $L^2(\mathbb{T})$. The proposition is taken directly from [15, Proposition 1.2.3]. For this proposition only, we will let $E_n(f) := \mathsf{dist}\ (L^2(\mathbb{T}); f, \mathbb{H}_n)$.

Proposition 2.15. *Let $r \geq 1$ be an integer, $f \in L^2(\mathbb{T})$. Then the following are equivalent:*

(a) *f has $r-1$ absolutely continuous derivatives and $f^{(r)} \in L^2(\mathbb{T})$.*

(b)$\sum_{k=1}^{\infty}\{k^r E_k(f)\}^2/k < \infty$.

(c)$\sum_{k=0}^{\infty}\{2^{kr} E_{2^k}(f)\}^2 < \infty$.

Such a stark connection between number of derivatives and degree of approximation by trigonometric polynomials does not arise for any other L^p space - instead we must use it as a motivating example to define a more useful smoothness class.

Definition 2.16. *Let $\gamma > 0$ and $\rho \in (0, \infty]$. Let* $\mathbf{a}$ *be a complex-valued sequence. Consider the seminorm*

$$\|\mathbf{a}\|_{\rho,\gamma} := \begin{cases} | \ \{2^{n\gamma}|a_n|\}_{n=0}^{\infty} \ |_{\rho}, & \text{if } 0 < \rho < \infty, \\ \sup\limits_{n \geq 0} 2^{n\gamma}|a_n|, & \text{if } \rho = \infty. \end{cases}$$

Then:

The Besov sequence space *is given by* $\mathsf{b}_{\rho,\gamma} = \{\mathbf{a} : \|\mathbf{a}\|_{\rho,\gamma} < \infty\}$.

The Besov approximation space *is given by* $B^p_{\rho,\gamma} = \{f \in X^p : \{E_{2^n,p}(f)\}_{n=0}^{\infty} \in \mathsf{b}_{\rho,\gamma}\}$*, with norm* $\|f\|_{p,\rho,\gamma} = \|f\|_p + \|\{E_{2^n,p}(f)\}_{n=0}^{\infty}\|_{\rho,\gamma}$.

When we speak of a Besov space, we will be referring to the Besov approximation space. For the univariate case, when necessary we will use $B^p_{\rho,\gamma}(\mathbb{T}) = \{f \in X^p(\mathbb{T}) : \{\mathsf{dist}\ (X^p(\mathbb{T}); f, \mathbb{H}_n)\}_{n=0}^{\infty} \in \mathsf{b}_{\rho,\gamma}\}$ As we saw in part (c) of Proposition 2.15, for $f \in L^2(\mathbb{T})$, the smoothness (in terms of the number r of derivatives) was classified (as both a sufficient and necessary condition) by $f \in B^2_{2,r}(\mathbb{T})$.

At first glance, the connection between Besov spaces and general smoothness of a function may not seem obvious, but in fact the development of tools for analyzing smoothness such as the K-functional and modulus of continuity in approximation theory leads to *equivalence theorems* connecting smoothness of a function with membership in an appropriate approximation space. For a modern treatment see Chapter 1 of [15] - we will state the equivalence theorem and terminology specifically as they apply here, for Besov spaces and approximation in X^p by trigonometric polynomials.

Definition 2.17. *Let $p \in [1, \infty]$, and $f \in X^p$. Then the K-functional of f between*

X^p and W_r^p is given by

$$\mathcal{K}(X^p, W_r^p; f, \delta) := \inf_{g \in W_r^p} \{\|f - g\|_p + \delta \|(-\Delta)^{r/2} g\|_p\}.$$

Note that when there exists a g for which the K-functional expression achieves its infimum, that g is the arg min and so solves the regularization problem implicitly posed, with $\|f - g\|_p$ the approximation term and $\|(-\Delta)^{r/2} g\|_p$ the loss functional, modulated by the parameter δ. A rigorous take on how this loss functional relates to the target function's smoothness, as well as the properties of the regularization problem's solution (which recovers the K-functional) is found for the case of L^2 in [7].

Note that when $g \in \mathbb{H}_{2^n}^q$, the Bernstein inequality Theorem 2.14 (b) gives that $\|(-\Delta)^{r/2} g\|_p \leq c 2^{nr} \|g\|_p$. This fact, along with the Favard inequality Theorem 2.14 (a), plays a critical role in understanding the equivalence theorem.

Theorem 2.18. *Let $p \in [1, \infty]$, $\rho \in (0, \infty]$, $\gamma > 0$, and $r > \gamma$ be an integer. Then*

$$\|f\|_{p,\rho,\gamma} \sim \|f\|_p + \left\| \{\mathcal{K}(X^p, W_r^p; f, 2^{-mr})\}_{m=0}^{\infty} \right\|_{\rho,\gamma}, \tag{2.4.22}$$

where the constants involved in $\sim$ may depend upon p, ρ, γ, and r.

While we will not prove this theorem here, as it is done in [15, Chapter 2, Theorem 4] in the full generality of the K-functional, we will mention that the complete proof requires an inequality facilitating the manipulation of sequences in Besov sequence spaces, known as a *discrete Hardy inequality*, which we later introduce and enlist in some proofs in Chapter 3.

This theorem states essentially that the Besov spaces characterize the smoothness of X^p functions in terms of K-functionals, and as we will see, the final piece of the puzzle is to observe that an arbitrary target function $f \in X^p$ lies in a Besov space exactly when the sequence of approximations by $\sigma_n^*(h,f)$ lie in the corresponding Besov sequence space; this completes the chain of equivalences from polynomial representation to smoothness classification.

2.5 Classification of Smoothness and Local Smoothness

As we showed in Section 2.4, given $f \in X^p$, the operators $\sigma_n^*(h,f)$ converge to f in the sense of L^p. This suggests that the σ_n^*'s act as partial sums of some series expansion for f - a correct assumption. Recall that h is some fixed low-pass filter, S-smooth with $S > q$.

Definition 2.19. *Let $p \in [1,\infty]$, and $f \in X^p$. Then the* ***band-pass operator*** *is given for non-negative integers n by*

$$\tau_0^*(h,f) = \sigma_1^*(h,f) \text{ and for } n \geq 1, \tau_n^*(h,f) = \sigma_{2^n}^*(h,f) - \sigma_{2^{n-1}}^*(h,f) \quad .$$

These band-pass operators form the aforementioned series expansion of $f \in X^p$, converging in the sense of L^p.

Theorem 2.20. *Let $p \in [1,\infty]$, and $f \in X^p$. Then we have that*

$$f = \sum_{j=0}^{\infty} \tau_j^*(h,f) \quad ,$$

with convergence of the infinite series in the sense of L^p.

Proof. Proving convergence is a simple affair, having the appropriate estimates on σ_n^*'s already. See that

$$\sum_{j=0}^{n} \tau_j^*(h,f) = \sigma_1^*(h,f) + \sum_{j=1}^{n} \left(\sigma_{2^j}^*(h,f) - \sigma_{2^{j-1}}^*(h,f)\right) = \sigma_{2^n}^*(h,f) \quad ,$$

so the partial sums give a telescoping series. Then we have by Proposition 2.11

$$\left\| f - \sum_{j=0}^{n} \tau_j^*(h,f) \right\|_p = \|f - \sigma_{2^n}^*(h,f)\|_p \leq c E_{2^{n-1},p}(f) \quad .$$

As $f \in X^p$, $E_{2^{n-1},p}(f) \to 0$ as $n \to \infty$, and we are done. $\square$

Now we are ready to state the classification of X^p function smoothness by trigonometric polynomial expansion. For the rest of the thesis we will use shorthand and refer to a sequence indexed by n, $\mathbf{a} = \{a_n\}_{n=0}^{\infty}$, as $\{a_n\}$, omitting the dependence on n whenever it is clear from context.

Theorem 2.21. *Let $p \in [1,\infty]$ and $f \in X^p$. Let $\rho \in (0,\infty]$ and $\gamma > 0$. Let $S > q$, and let h be an S-smooth low-pass filter. Then each of the following are equivalent:*

(a) $f \in B_{\rho,\gamma}^p$, *and so* $\{E_{2^n,p}(f)\} \in \mathsf{b}_{\rho,\gamma}$.

(b) $\{\|f - \sigma_{2^n}^*(h,f)\|_p\} \in \mathsf{b}_{\rho,\gamma}$.

(c) $\{\|\tau_n^*(h,f)\|_p\} \in \mathsf{b}_{\rho,\gamma}$.

This result reproduces part of Theorem 2.3.4 of [15], and we will prove it in greater generality as Theorem 3.14. However, the Besov space characterization of smoothness leaves more to be desired - two functions in the same approximation space may still have drastically different behavior. The following example comes from [15].

EXAMPLE 2.5.1. Consider the two functions defined on $\mathbb{T}$ by

$$f_1(x) = \sqrt{|\cos x|} \qquad \text{and} \qquad f_2(x) = \sum_{k=0}^{\infty} \frac{\cos(4^k x)}{2^k} \quad .$$

We see that $|f_1(x) - f_1(y)| \leq \sqrt{|\,|\cos x| - |\cos y|\,|} \leq \sqrt{|\cos x - \cos y|} \leq \sqrt{|x-y|}$ for $x, y \in \mathbb{T}$. The direct theorem of approximation theory (see [20]) then gives that $\mathsf{dist}\ (L^\infty(\mathbb{T}); f_1, \mathbb{H}_n) \leq cn^{-1/2}$, so that $f_1 \in B^\infty_{\infty,1/2}(\mathbb{T})$.

Next, we consider for m the largest integer such that $4^m \leq n$,

$$\begin{aligned} \mathsf{dist}\ (L^\infty(\mathbb{T}); f_2, \mathbb{H}_n) &\leq \mathsf{dist}\ (L^\infty(\mathbb{T}); f_2, \mathbb{H}_{4^m}) \\ &\leq \sum_{k=m+1}^{\infty} \left| \frac{\cos(4^k x)}{2^k} \right| \leq \sum_{k=m+1}^{\infty} 2^{-k} = 2^{-m} \leq cn^{-1/2} \quad . \end{aligned}$$

Likewise then, $f_2 \in B^\infty_{\infty,1/2}(\mathbb{T})$. Observing Figure 2.4 suggests a pitfall regarding this notion of smoothness - a singularity anywhere in the domain reduces the smoothness of the function on its entire domain, as defined with Besov spaces.

□

In the prior example, if one observed f_1 on a sufficiently small neighborhood of a point x_0 away from $\pm\pi/2$, then f_1 is much smoother than suggested by $B^\infty_{\infty,1/2}$ (in fact it is infinitely differentiable there) - this property clearly doesn't extend to f_2, the Weierstrass nowhere-differentiable function, even though it lies in the same approximation space. Let us now differentiate between the *global* approximation spaces we've been discussing, which characterize global smoothness, and *local* approximation spaces, which will characterize local smoothness.

- When $r > 0$, we will call a set of the form $\{\mathbf{x} : \|\mathbf{x} - \mathbf{x}_0\|_1 \leq r\}$ an r-box about $\mathbf{x}_0$.

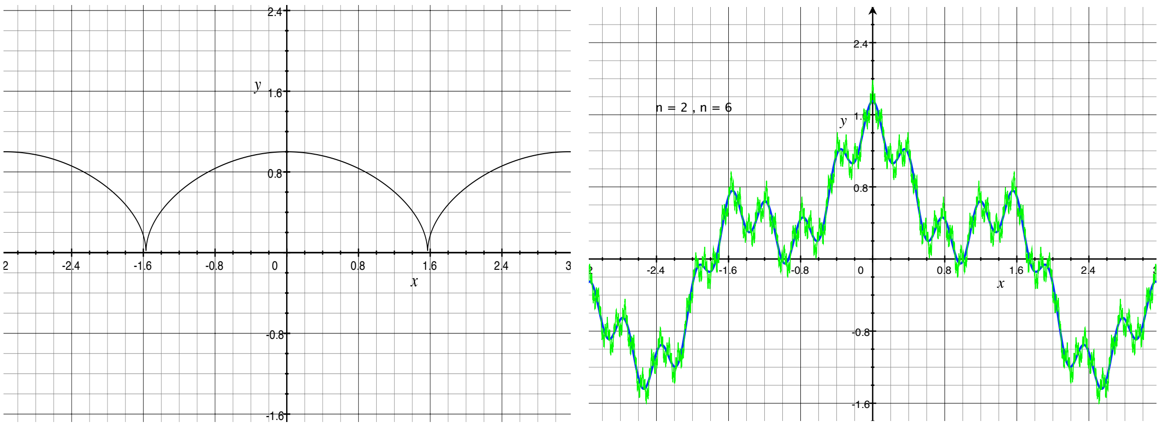

Figure 2.4: The graphs of $\sqrt{|\cos x|}$ and some partial sums (summing to the 2$^{\text{nd}}$ and 6$^{\text{th}}$ term) of the Weierstrass nowhere-differentiable function, both functions which lie in the approximation space $B^{\infty}_{\infty,1/2}(\mathbb{T})$. Evidently the local smoothness of these two functions is dramatically different away from the singularities of $\sqrt{|\cos x|}$.

- The space of functions $\varphi : \mathbb{T}^q \to \mathbb{C}$ where $D^{\mathbf{r}}\varphi$ is absolutely continuous for each $\mathbf{r} \in \mathbb{Z}^q_+$ is denoted C^∞.

- Given an r-box I, the space of C^∞ functions supported on I is denoted C^∞_I.

Definition 2.22. *Let $p \in [1,\infty]$, $\rho \in (0,\infty]$, $\gamma > 0$, and $\mathbf{x}_0 \in \mathbb{T}^q$. We will say that $f \in X^p$ is in the* local Besov approximation space $B^p_{\rho,\gamma}(\mathbf{x}_0)$ *when there exists an r-box I about $\mathbf{x}_0$ such that $\varphi \in C^\infty_I \implies f\varphi \in B^p_{\rho,\gamma}$.*

Any such r-box J about $\mathbf{x}_0$ for which we have $\varphi \in C^\infty_J \implies f\varphi \in B^p_{\rho,\gamma}$ will be called a (p,ρ,γ) Besov r-box *for f about $\mathbf{x}_0$.*

When we speak of local approximation, we are referring to the local Besov approximation spaces. $f \in B^p_{\rho,\gamma}(\mathbf{x}_0)$ is equivalent to the existence of a (p,ρ,γ) Besov r-box for f about $\mathbf{x}_0$. To finish our analysis of the prior example, while for any $x_0 \in \mathbb{T}$,

$f_2 \in B^{\infty}_{\infty,1/2}(x_0)$ and this is the best which can be hoped for, the same holds for f_1 only when $x_0 = \pm\pi/2$. For other choice of x_0, f_1 lies in $B^{\infty}_{\infty,\gamma}(x_0)$ for any $\gamma > 0$.

The final matter is to classify the local smoothness of a target function via the local approximation properties of its corresponding trigonometric polynomial expansion - with the following theorem in hand, we have achieved the desired wavelet-like representation using the τ^*_n operators.

Theorem 2.23. *Let $p \in [1,\infty]$ and $f \in X^p$. Let $\rho \in (0,\infty]$, $\gamma > 0$, and $S > q+\gamma$. Let h be an S-smooth low-pass filter, and $\mathbf{x}_0 \in \mathbb{T}^q$. Then we have the following:*

(a) *When I is a (p,ρ,γ) Besov r-box for f about $\mathbf{x}_0$, there exists a positive $r_0 < r$ and r_0-box J about $\mathbf{x}_0$ such that $\{\|f - \sigma^*_{2^n}(h,f)\|_{p,J}\} \in \mathsf{b}_{\rho,\gamma}$ and $\{\|\tau^*_n(h,f)\|_{p,J}\} \in \mathsf{b}_{\rho,\gamma}$.*

(b) *When for some r-box I about $\mathbf{x}_0$, either $\{\|f - \sigma^*_{2^n}(h,f)\|_{p,I}\} \in \mathsf{b}_{\rho,\gamma}$ or $\{\|\tau^*_n(h,f)\|_{p,I}\} \in \mathsf{b}_{\rho,\gamma}$, then I is a (p,ρ,γ) Besov r-box for f about $\mathbf{x}_0$.*

This is a mild rewriting of the first two statements of Theorem 2.4.3 of [15], done in our specific context, and once again rather than prove it here we will show a more general result in Theorem 3.16. We recall that the condition on h to achieve this characterization was $S > q+\gamma$, and that this means $\Phi_n(h,\mathbf{x})$ vanishes faster than $(n|\mathbf{x}|_1)^{-\gamma}$ - this is the localization criteria for achieving the desired description of local smoothness.

CHAPTER 3

Wavelet-Like Representation by Periodic Translation Networks

Here we will introduce the novel contribution of this thesis, comprising joint work with Professor Mhaskar. We begin by introducing periodic translation networks and an abstract notation for encapsulating scattered data and representing quadrature formulas. We proceed to develop an analogue of the theory of wavelet-like expansions in trigonometric polynomials, in the setting of periodic translation networks, and equipped to handle function approximation using scattered data.

Fix a positive integer q - as in Chapter 2, this is the dimension of our space. Consider a periodic *activation function* $\phi : \mathbb{T}^q \to \mathbb{C}$. A *periodic translation network* (PTN) with n neurons at the centers $\{\mathbf{x}_j\}_{j=1}^n$, with activation function ϕ, is defined as a function of the form

$$\mathbf{x} \mapsto \sum_{j=1}^{n} a_j \phi(\mathbf{x} - \mathbf{x}_j) \quad .$$

Hence a PTN is, as the name makes suggestive, a linear combination of translates of a periodic function. PTNs have been part of the literature (as in [17]) and have some useful qualities - for example, their network architecture allows hardware and software implementation using parallel computing. As noted in Chapter 1, a variety of minimal energy interpolation problems in the periodic setting give rise to PTNs.

The density of the class of PTNs for approximating functions in X^p was studied by Mhaskar and Micchelli in [18]. In particular, they proved the following theorem (cf. [18, eqn. (2.23), Proposition 2.1]):

Theorem 3.1. *Let $1 \leq p \leq \infty$, $\phi \in X^p$. The necessary and sufficient condition that the class of all PTNs with activation function ϕ be dense in X^p is that $\hat{\phi}(\mathbf{k}) \neq 0$ for any $\mathbf{k} \in \mathbb{Z}^q$.*

Accordingly, we make the following definition.

Definition 3.2. *A function $\phi \in L^1$ will be said to have **full frequency** if $\widehat{\phi}(\mathbf{k}) \neq 0$ for any $\mathbf{k} \in \mathbb{Z}^q$. For any ϕ with full-frequency we define for each $n > 0$ the positive quantity $m_n := \min_{|\mathbf{k}|_2 \leq n} |\widehat{\phi}(\mathbf{k})|$.*

The starting point of our theory is a representation of trigonometric polynomials as a convolution against the activation function. A great deal of early literature (for instance, see [27]) on this subject expresses some target function (in our case, polynomials) as an element of a reproducing kernel Hilbert space with ϕ as the reproducing kernel. This gives rise to the notion of native spaces. In our context, we borrow from this idea only to define a differential operator $\mathcal{D}_\phi$ with respect to a function ϕ with full frequency, by:

$$\widehat{f}(\mathbf{k}) = \widehat{\phi}(\mathbf{k})\ \widehat{\mathcal{D}_\phi f}(\mathbf{k}) \qquad \mathbf{k} \in \mathbb{Z}^q \quad ,$$

From Proposition 2.13, it is then prudent to consider the corresponding integral operator $\mathcal{I}_\phi$, a convolution with respect to the measure μ, against some function ϕ with full frequency:

$$\mathcal{I}_\phi g(\mu; \mathbf{x}) := \int_{\mathbb{T}^q} \phi(\mathbf{x} - \mathbf{y}) g(\mathbf{y}) \mathrm{d}\mu(\mathbf{y}) \quad .$$

Both of these operators are certainly well-defined for trigonometric polynomials as arguments, and map $\mathbb{H}_n^q$ into itself for each n. For the remainder of this chapter only, when $\mu = \mu_q^*$, we do not mention it and denote $\mathcal{I}_\phi g(\mathbf{x}) := \mathcal{I}_\phi g(\mu_q^*; \mathbf{x})$.

The punch line to the above is that these operators invert each other when evaluating a trigonometric polynomial with respect to the Lebesgue measure.

Proposition 3.3. *Let n be a positive integer, and let ϕ have full frequency. For $T \in \mathbb{H}_n^q$, we have*

$$T(\mathbf{x}) = (\phi * \mathcal{D}_\phi T)(\mathbf{x}) = \mathcal{I}_\phi \mathcal{D}_\phi T(\mathbf{x}) = \mathcal{D}_\phi \mathcal{I}_\phi T(\mathbf{x}).$$

Proof. When T is expressed as its Fourier series, notice that $\widehat{T}(\mathbf{k}) = 0$ when $|\mathbf{k}|_2 > n$:

$$\mathcal{I}_\phi T(\mathbf{x}) = \int_{\mathbb{T}^q} \phi(\mathbf{x} - \mathbf{y}) T(\mathbf{y}) \mathrm{d}\mu_q^*(\mathbf{y}) = \int_{\mathbb{T}^q} \phi(\mathbf{x} - \mathbf{y}) \sum_{\mathbf{k} \in \mathbb{Z}^q; |\mathbf{k}|_2 \le n} \widehat{T}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{y}} \mathrm{d}\mu_q^*(\mathbf{y}) \quad .$$

Redefining $\mathbf{y} \mapsto \mathbf{x} - \mathbf{y}$ and noting that the sum is finite,

$$\begin{aligned}
\mathcal{I}_\phi T(\mathbf{x}) &= \int_{\mathbb{T}^q} \phi(\mathbf{y}) \sum_{\mathbf{k} \in \mathbb{Z}^q; |\mathbf{k}|_2 \le n} \widehat{T}(\mathbf{k}) e^{i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \mathrm{d}\mu_q^*(\mathbf{y}) \\
&= \sum_{\mathbf{k} \in \mathbb{Z}^q; |\mathbf{k}|_2 \le n} \widehat{T}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \int_{\mathbb{T}^q} \phi(\mathbf{y}) e^{-i\mathbf{k}\cdot\mathbf{y}} \mathrm{d}\mu_q^*(\mathbf{y}) \\
&= \sum_{\mathbf{k} \in \mathbb{Z}^q; |\mathbf{k}|_2 \le n} \widehat{T}(\mathbf{k}) \widehat{\phi}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{x}} \quad .
\end{aligned}$$

Notice that expanding $\mathcal{I}_\phi T$ and equating Fourier coefficients (using uniqueness of the expansion) gives

$$\widehat{\mathcal{I}_\phi T}(\mathbf{k}) = \widehat{T}(\mathbf{k}) \widehat{\phi}(\mathbf{k}) \quad ,$$

which using the definition of $\mathcal{D}_\phi$ states that $T = \mathcal{D}_\phi \mathcal{I}_\phi T$; that is, the differential operator $\mathcal{D}_\phi$ inverts convolution by ϕ with respect to μ_q^*. Conversely, consider $\mathcal{I}_\phi \mathcal{D}_\phi T$.

For the sake of being concise, we note that by Proposition 2.13, for each $\mathbf{k} \in \mathbb{Z}^q$,

$$\widehat{\mathcal{I}_\phi \mathcal{D}_\phi T}(\mathbf{k}) = \widehat{\phi}(\mathbf{k})\widehat{\mathcal{D}_\phi T}(\mathbf{k}) = \widehat{T}(\mathbf{k}) \quad ,$$

invoking the definition of $\mathcal{D}_\phi T$. As T and $\mathcal{I}_\phi \mathcal{D}_\phi T$ are polynomials, $\mathcal{I}_\phi \mathcal{D}_\phi T = T$ and we are done. $\square$

The significance of this proposition is that for any trigonometric polynomial T, convolving $\mathcal{D}_\phi T$ against ϕ reproduces T. Hence having solved a problem of approximation in the setting of trigonometric polynomials, these approximations can be reproduced as desired. The convolution itself however opens a new door for us.

3.1 Discretizing Integrals

In order to produce a PTN from the polynomial-reproducing relation, we must discretize the integral - that is, we must be able to represent this integral against the Lebesgue measure as a sum against finitely many centers. Because this type of discretization will appear again shortly in connection with function approximation from scattered data, we will use a more abstract notation for convenience.

Suppose $f : \mathbb{R}^q \to \mathbb{C}$, and for each $n = 0, 1, 2, \ldots$, the value of f is known on N_n inputs $\mathbf{x}_{kn}$, $k = 1, 2, \ldots, N_n$. A *discretization* of the integral of f of order n is an assignment of real-valued weights $\{w_{kn}\}_{k=1}^{N_n}$ such that $\int_{\mathbb{T}^q} f \mathrm{d}\mu_q^* \approx \sum_{k=1}^{N_n} w_{kn} f(\mathbf{x}_{kn})$. The term 'order' as used here doesn't come intrinsically from the definition, but instead inherits whatever meaning the index n has; eg. when discretizing a sequence (T_n) of polynomials, where $T_n \in \mathbb{H}_n^q$ for each n. Given such a discretization associated with a sequence, it will be useful for reasons we will discuss shortly, to define for each

n a corresponding finitely supported measure

$$\nu_n(\mathbf{x}) = \sum_{k=1}^{N_n} w_{kn}\delta(\mathbf{x} - \mathbf{x}_{kn}) \quad , \tag{3.1.1}$$

where $\delta(\mathbf{x})$ is the Dirac delta distribution restricted to $\mathbb{T}^q$. In this case the finite support is clear, as ν_n is supported on $\{\mathbf{x}_{kn}\}_{k=1}^{N_n}$. Then we may neatly represent the discretization as an integral against the finitely supported measure, since

$$\int_{\mathbb{T}^q} f \mathrm{d}\nu_n = \sum_{k=1}^{N_n} w_{kn} f(\mathbf{x}_{kn}) \quad .$$

To fully enlist the utility of the abstract measure notation, we require two more concepts related to discretization, which we take from [6]. With regard to notation, ν will always refer to a finitely supported measure which discretizes some integral expression, and μ will be used when discussing general situations in which either a finite measure, or the usual Lebesgue measure may apply. In fact, this theory presumes even greater generality, but the applications we have studied enlist only these two cases.

We recall that the total variation measure of any signed measure μ is defined by

$$|\mu|(\mathcal{U}) := \sup \sum_{i=1}^{\infty} |\mu(U_i)|, \qquad \mathcal{U} \subset \mathbb{T}^q \quad ,$$

where the supremum is taken over all countable partitions $\{U_i\}$ into measurable sets of $\mathcal{U}$. In the case when $\mu = \nu_n$ as in (3.1.1), one can easily deduce that $|\nu_n|(\mathbb{T}^q) = \sum_{k=1}^{N_n} |w_{kn}|$. The Hahn-Jordan Decomposition (as in [26]) states that for any signed measure μ defined on a σ-algebra Σ, there exist positive measures μ_+, μ_- defined on a partition of Σ such that $\mu_+ - \mu_- = \mu$ and $\mu_+ + \mu_- = |\mu|$.

In order to discuss approximation in the setting of more general measures, first we must generalize L^p norms to account for various measures. For $p \in [1, \infty]$, μ a (possibly signed) measure on $\mathbb{R}^q$, $A \subset \mathbb{R}^q$ μ-measurable, and $f : A \to \mathbb{C}$ μ-measurable, the L^p norm of f with respect to μ is given by

$$\|f\|_{\mu;p,A} := \begin{cases} \left\{ \int_A |f(\mathbf{x})|^p \mathrm{d}|\mu|(\mathbf{x}) \right\}^{1/p}, & \text{if } 1 \le p < \infty \quad , \\ |\mu| - \operatorname*{ess\,sup}_{\mathbf{x} \in A} |f(\mathbf{x})|, & \text{if } p = \infty. \end{cases} \tag{3.1.2}$$

This is a norm when we consider f to be the equivalence class of all functions equal to f $|\mu|$-almost everywhere, and we denote by $L^p(\mu; A)$ the linear space of functions with finite L^p norm with respect to μ. Note that the definition of almost everywhere involves the measure; hence we specify $|\mu| - \text{ess sup}$ to accentuate this reliance when not using the Lebesgue measure. As usual, when $A = \mathbb{T}^q$ we write $L^p(\mu) := L^p(\mu; \mathbb{T}^q)$ and $\|f\|_{\mu;p} := \|f\|_{\mu;p,\mathbb{T}^q}$.

The concepts we would like to extend to abstract measures are discretization of integrals and good approximation of trigonometric polynomials. Together a measure which has these qualities allows our analysis of expansions in translates of functions to have a succinct form which relies directly on the properties of multivariate trigonometric polynomial expansions already discussed in Chapter 2.

Definition 3.4. (a) *A (possibly signed) measure μ is called a* ***quadrature measure*** *of order n when*

$$\int_{\mathbb{T}^q} T \mathrm{d}\mu = \int_{\mathbb{T}^q} T \mathrm{d}\mu_q^* \qquad \forall T \in \mathbb{H}_n^q \quad . \tag{3.1.3}$$

(b) *A (possibly signed) measure μ is called a* ***Marcinkiewicz–Zygmund measure****, or M-Z measure, of order n when the M-Z inequality below is satisfied, with $c(n, \mu)$ a* *constant independent of T:*

$$\int_{\mathbb{T}^q} |T|\mathrm{d}|\mu| = \|T\|_{\mu;1} \leq c(n,\mu)\|T\|_1 \qquad \forall T \in \mathbb{H}_n^q \quad . \tag{3.1.4}$$

When μ is a finitely supported quadrature measure, the summation expression which gives the Lebesgue integral is called a *quadrature formula* in the literature, and can be thought of as a discretization where the support of the measure may represent the known information of f, or the desired centers for a network expansion. It can easily be shown that the smallest c that works in (3.1.4) which may be chosen for a measure μ of order n is a norm on the space of Radon measures, which we denote by $|\!|\!|\mu|\!|\!|_n$, the **M-Z norm**. Radon measures are complete, regular Borel measures - to avoid bringing in technicalities from measure theory, we refer the reader to the details in [26]. Without using measure notation, this norm would have to be formulated directly in terms of the weights w_{kn} and centers $\mathbf{x}_{kn}$, and it is not at all clear what specific properties of this discretization the M-Z inequality constant would depend on, or even if it would be specific values or their behavior in aggregate. Considerations like this are a significant motivation for our choice of notation.

A concrete example of an M-Z quadrature measure which satisfies both conditions for any order n is of course the Lebesgue measure on $\mathbb{T}^q$. A more interesting choice is the measure generated by the quadrature formula for integrating a univariate trigonometric polynomial over equidistant nodes. We found this example in [15] and slightly modified it.

Proposition 3.5. *Let n be a positive integer. Consider the case $q = 1$, centers $x_{jn} = \frac{2\pi j}{n+1} - \pi$, and weights $w_{jn} = \frac{1}{n+1}$. Then the measure $\nu_n(x) = \sum_{j=0}^{n} w_{jn}\delta(x - x_{jn})$*

is an M-Z quadrature measure of order n.

Proof. Let $T \in \mathbb{H}_n^1 = \mathbb{H}_n$. We express T as $\sum_{|k|\le n} a_k e^{ik\circ}$, and recall that from orthonormality we have $\int_{\mathbb{T}} e^{i\ell x} \mathrm{d}\mu_1^*(x) = \delta_{\ell,0}$. Hence $\int_{\mathbb{T}} T \mathrm{d}\mu_1^* = a_0$. The quadrature formula may be written

$$\int_{\mathbb{T}} e^{i\ell x} \mathrm{d}\nu_n(x) = \frac{1}{n+1} \sum_{j=0}^{n} \exp\left(\frac{2\pi i \ell j}{n+1} - \pi i \ell\right) = \frac{(-1)^\ell}{n+1} \sum_{j=0}^{n} \exp\left(\frac{2\pi i \ell j}{n+1}\right) \quad .$$

Should $n+1$ be a factor of ℓ, each term in the summation is 1 and the formula evaluates to $(-1)^\ell$. Else, this is a geometric series and evaluates to

$$\frac{1 - \exp(2\pi i \ell)}{1 - \exp\left(\frac{2\pi i \ell}{n+1}\right)} = 0 \quad ,$$

since the denominator doesn't vanish. So, because $T \in \mathbb{H}_n$, the only term of T for which the degree k satisfies $n+1 \mid k$ is $k = 0$, and we conclude

$$\int_{\mathbb{T}} T(x) \mathrm{d}\nu_n(x) = \sum_{j=0}^{n} \frac{1}{n+1} T(x_{jn}) = \sum_{j=0}^{n} \sum_{|k|\le n} \frac{a_k}{n+1} e^{x_{jn}} = \sum_{|k|\le n} a_k (-1)^k \delta_{k,0} = a_0 \quad .$$

Hence ν_n is a quadrature measure of order n.

Now consider the error term expression

$$\left| \frac{1}{n+1} \sum_{j=0}^{n} |T(x_{jn})| - \|T\|_1 \right| = \left| \frac{1}{n+1} \sum_{j=0}^{n} |T(x_{jn})| - \int_{\mathbb{T}} |T(t)| \mathrm{d}\mu_1^*(t) \right| \quad .$$

For this proof only, let $I_{jn} := [x_{jn}, x_{j+1\,n}]$; $\mathbb{T}$ is partitioned into $n+1$ such intervals. Notice that $\mu_1^*(I_{jn}) = \frac{1}{n+1}$. So, the above expression can be written

$$\begin{aligned} \left| \sum_{j=0}^{n} \int_{I_{jn}} \left(|T(x_{jn})| - |T(t)|\right) \mathrm{d}\mu_1^*(t) \right| &\le \sum_{j=0}^{n} \int_{I_{jn}} \Big| |T(x_{jn})| - |T(t)| \Big| \mathrm{d}\mu_1^*(t) \\ &\le \sum_{j=0}^{n} \int_{I_{jn}} |T(x_{jn}) - T(t)| \, \mathrm{d}\mu_1^*(t) \quad . \end{aligned}$$

Note that for $t \in I_{jn}$, $|T(x_{jn}) - T(t)| \leq \left|\int_t^{x_{jn}} T(u)\mathrm{d}u\right| \leq \int_{I_{jn}} |T'(u)|\mathrm{d}u$ with $\mathrm{d}u = 2\pi\mathrm{d}\mu_1^*(u)$ here meaning the Lebesgue measure on $\mathbb{R}$. Using the Bernstein inequality (2.4.20), we then conclude

$$\begin{aligned}
\sum_{j=0}^{n} \int_{I_{jn}} |T(x_{jn}) - T(t)|\, \mathrm{d}\mu_1^*(t) &\leq \sum_{j=0}^{n} \int_{I_{jn}} \int_{I_{jn}} |T'(u)|\, \mathrm{d}u\, \mathrm{d}\mu_1^*(t) \\
&= \frac{2\pi}{n+1} \int_{\mathbb{T}} |T'(u)|\, \mathrm{d}\mu_1^*(u) \\
&\leq 2\pi \frac{n}{n+1} \int_{\mathbb{T}} |T(u)| \mathrm{d}\mu_1^*(u) \quad .
\end{aligned}$$

Hence

$$\|T\|_{\nu_n;1} = \frac{1}{n+1} \sum_{j=0}^{n} |T(x_{jn})| \leq \left| \frac{1}{n+1} \sum_{j=0}^{n} |T(x_{jn})| - \|T\|_1 \right| + \|T\|_1 \leq (2\pi + 1)\|T\|_1 \quad ,$$

so ν_n is an M-Z measure of order n. □

Remark 3.6. *It is interesting to note that in [6], Filbir and Mhaskar show that a positive quadrature measure is automatically an M-Z measure, and that given an M-Z measure of order n and some positive constant k, this measure is also an M-Z measure of order kn with equivalent M-Z norm.*

Critical for applications, Mhaskar showed in [13] that given any sufficiently well-distributed, finite set of data in $\mathbb{T}^q$, an M-Z quadrature measure can be constructed supported on the data. How well some finite set of data $\mathcal{C}$ is distributed on $\mathbb{T}^q$ is measured by the *density content* $\delta(\mathcal{C}) := \max_{\mathbf{x} \in \mathbb{T}^q} \min_{\mathbf{y} \in \mathcal{C}} \|\mathbf{x} - \mathbf{y}\|_2$.

As we will rely on this result whenever we deal with scattered data, the result is reproduced here.

Theorem 3.7. *Let $\mathcal{C}_0$ be a set of distinct points in $\mathbb{T}^q$ and $n \geq 1$ be an integer such that*

$\delta_{\mathcal{C}_0} < \pi/(2 \cdot 3^{q+3} n)$. *Then there exist numbers* $\{w_{\boldsymbol{\xi}}\}_{\boldsymbol{\xi} \in \mathcal{C}_0}$ *such that*

$$|w_{\boldsymbol{\xi}}| \leq \frac{c}{n^q}, \qquad \boldsymbol{\xi} \in \mathcal{C}_0, \tag{3.1.5}$$

and for every $T \in \mathbb{H}_n^q$,

$$\int_{\mathbb{T}^q} T(\mathbf{t}) d\mu_q^*(\mathbf{t}) = \sum_{\boldsymbol{\xi} \in \mathcal{C}_0} w_{\boldsymbol{\xi}} T(\boldsymbol{\xi}). \tag{3.1.6}$$

In particular,

$$\sum_{\boldsymbol{\xi} \in \mathcal{C}_0} |w_{\boldsymbol{\xi}}||T(\boldsymbol{\xi})| \leq \int_{\mathbb{T}^q} |T(\mathbf{t})| d\mu_q^*(\mathbf{t}) \quad . \tag{3.1.7}$$

We may also choose $w_{\boldsymbol{\xi}}$ *to be non-negative instead of requiring (3.1.5).*

The estimate (3.1.7) is a simple consequence of [13, Theorem 3.3.2] and the remaining statements of the above theorem which are [13, Theorem 3.3.1]. Efficient numerical techniques have been developed and utilized in [11] to construct such quadrature formulas from sufficiently dense sets of data.

Hence the only restriction we place on finite data for use in approximation as described here, is that as the amount of data increases it grows dense on $\mathbb{T}^q$. The fundamental utility of M-Z quadrature measures here is that they retain good approximation properties for some polynomial-reproducing convolutions.

3.2 Approximation on Scattered Data

In order to distinguish convolution against polynomial kernels which we discuss here, from the convolution against full-frequency functions ϕ called $\mathcal{I}_\phi$ defined in Section 1, we introduced a new notation for clarity.

Definition 3.8. *Let μ be a Borel measure on $\mathbb{T}^q$. Let $f, g \in L^1(\mu)$. Then the convolution of f against g with respect to μ is defined by*

$$(f *_\mu g)(\mathbf{x}) = \int_{\mathbb{T}^q} f(\mathbf{y}) g(\mathbf{x} - \mathbf{y}) \mathrm{d}\mu(\mathbf{y}) \quad .$$

Now to include scattered data approximation into our discussion, we will need to pause and develop discretized summability operators, with respect to (potentially finitely supported) M-Z measures, and recover analogous localization results in order for the same theorems applying to σ_n^* to pass along trivially.

Definition 3.9. *Let n be a positive integer, let μ be a Borel measure on $\mathbb{T}^q$. Let $p \in [1, \infty]$, and $f \in X^p(\mu)$. Let h be a low-pass filter. Then the summability operator with respect to μ is given by*

$$\sigma_n(\mu; h, f)(\mathbf{x}) := (f *_\mu \Phi_n(h))(\mathbf{x}) = \sum_{|\mathbf{k}|_2 \leq n} h\left(\frac{|\mathbf{k}|_2}{n}\right) e^{i\mathbf{k}\cdot\mathbf{x}} \int_{\mathbb{T}^q} f(\mathbf{t}) e^{-i\mathbf{k}\cdot\mathbf{t}} \mathrm{d}\mu(\mathbf{t}) \quad .$$

We note but do not pursue further, that the expression $\int_{\mathbb{T}^q} f(\mathbf{t}) e^{-i\mathbf{k}\cdot\mathbf{t}} \mathrm{d}\mu(\mathbf{t})$ may be seen as a generalized Fourier coefficient with respect to μ. In the case $\mu = \mu_q^*$, we recover $\sigma_n(\mu_q^*; h, f) = \sigma_n^*(h, f)$. Further, we fix some S-smooth low-pass filter h, with $S > q$, for the remainder of the chapter and will omit it from notation, as its choice does not affect the analysis. At times we may demand $S > q + \gamma$ for some $\gamma > 0$; presume the selection of a corresponding S-smooth low-pass filter h, and again omit it from notation. Such a choice is always possible (e.g. h_∞ from Example 2.3.2), and may affect only the particular constants in formulas.

Theorem 3.10. *Let n be a positive integer, and let μ be an M-Z quadrature measure of order $3/2\ n$. Let $p \in [1, \infty]$.*

(a) *For* $T \in \mathbb{H}^q_{n/2}$, $\sigma_n(\mu; T) = T$.

(b) *We have for* $f \in X^p(\mu)$,

$$\|\sigma_n(\mu; f)\|_p \le c|\!|\!|\mu|\!|\!|_n^{1-1/p}\|f\|_{\mu;p} \quad . \tag{3.2.8}$$

Consequently, if $f \in X^\infty$, *then* $\|\sigma_n(\mu; f)\|_\infty \le c|\!|\!|\mu|\!|\!|_n\|f\|_\infty$, *and further*

$$E_{n,\infty}(f) \le \|f - \sigma_n(\mu; f)\|_\infty \le c|\!|\!|\mu|\!|\!|_n E_{n/2,\infty}(f) \quad . \tag{3.2.9}$$

(c) *If* $f \in L^1(\mu)$ *is supported on a compact set* K, *and* V *is an open set with* $K \subset V$, *then*

$$\|\sigma_n(\mu; f)\|_{\infty, \mathbb{T}^q \setminus V} \le c\|f\|_{\mu;1} n^{q-S}, \tag{3.2.10}$$

where c *may depend upon* K *and* V *in addition to* S *and* h.

The statement and proof of Theorem 3.10 follows the corresponding statements and proofs in [16]. First, we state a necessary generalization of the Young inequality to convolutions with respect to an arbitrary Borel measure on $\mathbb{T}^q$. The result may be proved as a consequence of the statement [15, Proposition 3.1.2] and the translation invariance of the Lebesgue measure; we do not show the proof here.

Lemma 3.11. *Let* μ *be a (possibly signed) Borel measure on* $\mathbb{T}^q$. *Let* f *be bounded and* μ*-measurable, and* $g \in L^1(\mu)$. *Then given* $p \in [1, \infty]$, *if* $f \in L^p(\mu)$, *we have that* $f *_\mu g \in L^p$ *and*

$$\|f *_\mu g\|_p \le \operatorname*{ess\,sup}_{\mathbf{x} \in \mathbb{T}^q} \|g(\mathbf{x} - \circ)\|_{\mu;1}^{1-1/p}\|g\|_1^{1/p}\|f\|_{\mu;p} \quad . \tag{3.2.11}$$

In particular, when μ *is an M-Z measure of order* N, *then for* $T \in \mathbb{H}^q_N$, *the above inequality may be stated*

$$\|f *_\mu T\|_p \le \sup_{\mathbf{x} \in \mathbb{T}^q} \|T(\mathbf{x} - \circ)\|_{\mu;1}^{1-1/p}\|T\|_1^{1/p}\|f\|_{\mu;p} \le |\!|\!|\mu|\!|\!|_N^{1-1/p}\|f\|_{\mu;p}\|T\|_1 \quad . \tag{3.2.12}$$

Note that we will often require the M-Z norm of an M-Z measure μ whose order has the form $3/2\ n$ or $3 \cdot 2^{n-1}$. In this case, as mentioned in Remark 3.6, μ is also an M-Z measure of order n and 2^n respectively, so its M-Z norm may be given equivalently by $|||\mu|||_n$ and $|||\mu|||_{2^n}$, respectively.

Now we may prove Theorem 3.10.

Proof. If $T \in \mathbb{H}_{n/2}$, $\mathbf{x} \in \mathbb{T}^q$, then $\Phi_n(\mathbf{x} - \circ)T(\circ) \in \mathbb{H}^q_{3n/2}$. Because μ is a quadrature measure of order $3n/2$, using the quadrature formula (3.1.3) we may easily verify by the definition of Φ_n, that

$$\int_{\mathbb{T}^q} \Phi_n(\mathbf{x} - \mathbf{t})T(\mathbf{t})\mathrm{d}\mu(\mathbf{t}) = \int_{\mathbb{T}^q} \Phi_n(\mathbf{x} - \mathbf{t})T(\mathbf{t})\mathrm{d}\mu_q^*(\mathbf{t}) = T(\mathbf{x}) \quad .$$

This proves part (a).

Because μ is an M-Z measure of order $3n/2 > n$, we may employ the general Young inequality in the form (3.2.12), along with Theorem 2.6, to show estimate (3.2.8):

$$\|\sigma_n(\mu; f)\|_p = \|f *_\mu \Phi_n\|_p \le |||\mu|||_n^{1-1/p} \|f\|_{\mu;p} \|\Phi_n\|_1 \le c|||\mu|||_n^{1-1/p} \|f\|_{\mu;p} \quad .$$

For arbitrary $T \in \mathbb{H}^q_{n/2}$, then part (a) and (3.2.8) imply that

$$E_{n,p}(f) \le \|f - \sigma_n(\mu; f)\|_p = \|f - T - \sigma_n(\mu; f - T)\|_p \le \|f - T\|_p + c|||\mu|||_n^{1-1/p} \|f - T\|_{\mu;p} \quad .$$

When $p = \infty$, we know that the essential supremum of a continuous function is actually the supremum; hence for $f \in X^\infty$, f is equal a.e. to a continuous function, and

$$\|f\|_\infty = \|f\|_{\mu;\infty} = \sup_{\mathbf{x} \in \mathbb{T}^q} |f(\mathbf{x})| \quad ,$$

and so $E_{n,\infty}(f) \leq \|f - \sigma_n(\mu; f)\|_\infty \leq c|\!|\!|\mu|\!|\!|_n \|f - T\|_\infty$. As T was arbitrary in $\mathbb{H}_{n/2}$, using the same argument as in Proposition 2.11 we conclude that $\|f - \sigma_n(\mu; f)\|_\infty \leq c|\!|\!|\mu|\!|\!|_n E_{n/2,\infty}(f)$. This proves (3.2.9).

Now let K be a compact subset of $\mathbb{T}^q$, $f \in L^1(\mu)$, and $\mathrm{supp}(f) \subset K$. Then if V is an open set with $K \subset V$, from Theorem 2.6 and (3.2.11) we have

$$\begin{aligned} \|\sigma_n(\mu; f)\|_{\infty,\mathbb{T}^q\setminus V} &= \sup_{\mathbf{x}\in\mathbb{T}^q\setminus V} |(f *_\mu \Phi_n)(\mathbf{x})| \\ &\leq \sup_{\substack{\mathbf{x}\in T^q\setminus V \\ \mathbf{t}\in K}} |\Phi_n(\mathbf{x} - \mathbf{t})| \; \|f\|_{\mu,1} \leq c\|f\|_{\mu;1} n^{q-S} \quad , \end{aligned}$$

noting in the last step that because K and $\mathbb{T}^q \setminus V$ are compact,

$$\min_{\substack{\mathbf{x}\in\mathbb{T}^q\setminus V \\ \mathbf{t}\in K}} |\mathbf{x} - \mathbf{t}|_1 \geq \epsilon(K, V) > 0 \quad ,$$

which is incorporated in the constant c. This shows (3.2.10). □

The approximation results for the rest of the chapter will each come in two forms, with starred operators representing general X^p approximation, and the general operators additionally capable of modeling continuous functions on scattered data in X^∞. For example, measures μ_n for positive integers n as in the above theorem may encode nested sets of scattered data sites $\mathcal{C}_n$. Should these sets satisfy the density conditions according to Theorem 3.7 to form, in each case, an M-Z quadrature measure of order n, Theorem 3.10 (b) then shows that the summability operators of f taken with respect to these measures will converge in the supremum norm to f.

Additionally, the results regarding expansion, smoothness classes, and local smoothness classes similarly carry over to the case of operators on general M-Z quadrature measures, giving a characterization in terms of Besov spaces and local Besov spaces.

First, we will need analogous band-pass operators. We say a sequence of measures $\boldsymbol{\mu} = \{\mu_n\}$ has nested support when

$$m < n \Longrightarrow \mathrm{supp}(\mu_m) \subset \mathrm{supp}(\mu_n) \quad .$$

When dealing with scattered data, measures with nested support represent accruing data, such that increasing n may be interpreted as keeping the data from an earlier experiment and potentially improving it with more data from a later experiment.

Definition 3.12. *Let $\boldsymbol{\mu}$ be a sequence of Borel measures on $\mathbb{T}^q$ with nested support. Let $p \in [1, \infty]$, and $f \in X^p(\mu)$. Then the band-pass operator with respect to $\boldsymbol{\mu}$ is given for each non-negative integer n as*

$$\tau_0(\boldsymbol{\mu}; f) = \sigma_1(\mu_0; f) \text{ and for } n \geq 1, \tau_n(\boldsymbol{\mu}; f) = \sigma_{2^n}(\mu_n; f) - \sigma_{2^{n-1}}(\mu_{n-1}; f) \quad .$$

We will use $\boldsymbol{\mu}^*_{\boldsymbol{q}}$ to denote the constant sequence of Lebesgue measures. Then $\tau_n(\boldsymbol{\mu}^*_{\boldsymbol{q}}; f) = \tau^*_n(f)$.

Proposition 3.13. *Let $f \in X^\infty$, and let $\boldsymbol{\mu} = \{\mu_n\}$ be a sequence of M-Z quadrature measures with nested support, each of order $3 \cdot 2^{n-1}$. Let $|\!|\!|\mu_n|\!|\!|_{2^n}$ be uniformly bounded, such that $|\!|\!|\mu_n|\!|\!|_{2^n} \leq c$ for each n. Then we have*

$$f = \sum_{j=0}^{\infty} \tau_j(\boldsymbol{\mu}; f) \quad ,$$

with convergence in the sense of L^∞.

Proof. This result is the discretized counterpart to Theorem 2.20. The partial sums to the series above form a telescoping series:

$$\sum_{j=0}^{n} \tau_j(\boldsymbol{\mu}; f) = \sigma_1(\mu_0; f) + \sum_{j=1}^{n} \left(\sigma_{2^j}(\mu_j; f) - \sigma_{2^{j-1}}(\mu_{j-1}; f)\right) = \sigma_{2^n}(\mu_n; f) \quad .$$

Then by Theorem 3.10 (b) and the uniform boundedness of $|||\mu_n|||_{2^n}$, we have

$$\left\| f - \sum_{j=0}^{n} \tau_j(\boldsymbol{\mu}; f) \right\|_\infty = \|f - \sigma_{2^n}(\mu_n; f)\|_\infty \leq c|||\mu_n|||_{2^n} E_{2^{n-1},\infty}(f) \leq cE_{2^{n-1},\infty}(f) \quad ,$$

where c does not depend on n. As $f \in X^\infty$, $E_{2^{n-1},\infty}(f) \to 0$ as $n \to \infty$, and we are done. $\square$

Whenever we approximate and measure the error in such expansions in an L^p norm, should $p \neq \infty$ and so f may not be equal a.e. to a continuous function, attempting to approximate f using scattered data is a hopeless endeavor - any finite set of data has measure zero and as such gives us effectively no information about the class of functions we seek at all. The problem as such is ill-posed. Using X^∞ however, addresses these pitfalls and we may use the full strength of the theory - this is the contextual explanation for why the following proofs and results carefully describe when general M-Z measures may be used in place of the Lebesgue measure.

Theorem 3.14. *Let $\boldsymbol{\mu} = \{\mu_n\}$ be a sequence of M-Z quadrature measures with nested support, each of order $3 \cdot 2^{n-1}$. Let $|||\mu_n|||_{2^n}$ be uniformly bounded, such that $|||\mu_n|||_{2^n} \leq c$ for each n. Then, given $\gamma > 0$, $\rho \in (0, \infty]$, $p \in [1, \infty]$, and $f \in X^p$, the following statements are equivalent:*

(a) $f \in B^p_{\rho,\gamma}$, *and so* $\{E_{2^n,p}(f)\} \in \mathsf{b}_{\rho,\gamma}$.

(b) $\left\{\|f - \sigma_{2^n}(\mu_q^*; f)\|_p\right\} \in \mathsf{b}_{\rho,\gamma}$.

(c) $\left\{\|\tau_n(\boldsymbol{\mu}_{\boldsymbol{q}}^*; f)\|_p\right\} \in \mathsf{b}_{\rho,\gamma}$.

Further, for $p = \infty$ the same equivalences hold with $\boldsymbol{\mu}$ in place of $\boldsymbol{\mu}_{\boldsymbol{q}}^*$, and μ_n replacing μ_q^*.

The equivalences as written are effectively the statement of Theorem 2.21 - the novelty here is the last statement, for $p = \infty$ and scattered data approximation. In order to complete the proof, we will require, as mentioned in Chapter 2, a discrete Hardy inequality which we reproduce here, as given (along with a proof) in [15, Chapter 1, Lemma 2.1]. We will not reproduce the proof.

Lemma 3.15. *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be sequences of nonnegative numbers,* $\rho, \gamma, \upsilon > 0$. *Suppose the following conditions holds:*

$$b_k \le c_0 \left(\sum_{j=k}^{\infty} a_j^{\upsilon} \right)^{1/\upsilon}, \qquad k = 0, 1, \cdots, \tag{3.2.13}$$

Then

$$\|\mathbf{b}\|_{\rho,\gamma} \le c c_0 \|\mathbf{a}\|_{\rho,\gamma}. \tag{3.2.14}$$

Now we may prove Theorem 3.14.

Proof. We will show (a) $\iff$ (b) $\iff$ (c).

In view of (2.3.18), we have

$$\|f - \sigma_{2^n}(\mu_q^*; f)\|_p \le c E_{2^{n-1},p}(f) \quad ,$$

from which we conclude

$$\|\{\|f - \sigma_{2^n}(\mu_q^*; f)\|_p\}\|_{\rho,\gamma} \le c \|\{E_{2^n,p}(f)\}\|_{\rho,\gamma} \quad .$$

So, (a) $\Longrightarrow$ (b). Likewise by (2.3.18) $E_{2^n,p}(f) \le \|f - \sigma_{2^n}(\mu_q^*; f)\|_p$, so (b) $\Longrightarrow$ (a).

Notice that

$$\begin{aligned} \|\tau_n(\boldsymbol{\mu}_{\boldsymbol{q}}^*; f)\|_p &= \|\sigma_{2^n}(\mu_q^*; f) - \sigma_{2^{n-1}}(\mu_q^*; f)\|_p \\ &\le \|f - \sigma_{2^n}(\mu_q^*; f)\|_p + \|f - \sigma_{2^{n-1}}(\mu_q^*; f)\|_p \quad , \end{aligned}$$

from which (b) $\implies$ (c). In the reverse direction, from Proposition 3.13 we have

$$\|f - \sigma_{2^n}(\mu_q^*; f)\|_p = \left\| \sum_{j=0}^{\infty} \tau_j(\boldsymbol{\mu}_{\boldsymbol{q}}^*; f) - \sum_{j=0}^{n} \tau_j(\boldsymbol{\mu}_{\boldsymbol{q}}^*; f) \right\|_p \le \sum_{j=n+1}^{\infty} \|\tau_j(\boldsymbol{\mu}_{\boldsymbol{q}}^*; f)\|_p.$$

Then the Hardy inequality (3.2.13) with $\upsilon = 1$ yields that (c) $\implies$ (b).

When $p = \infty$, we may directly replace each instance of $\sigma_{2^n}(\mu_q^*; f)$ and $\tau_n(\boldsymbol{\mu}_{\boldsymbol{q}}^*; f)$ in the proof and theorem with $\sigma_{2^n}(\mu_n; f)$ and $\tau_n(\boldsymbol{\mu}; f)$ respectively, and due to Theorem 3.10 and the uniform boundedness of $|||\mu_n|||_{2^n}$, with (3.2.9) in place of (2.3.18), each statement will hold as desired. $\square$

The characterization of local Besov spaces proceeds similarly, and implicitly requires the high localization in the form of estimate (3.2.10).

Theorem 3.16. *Assume the conditions of Theorem* 3.14*, and additionally let* $S > q + \gamma$*, and let* $\mathbf{x}_0 \in \mathbb{T}^q$*. Then we have the following:*

(a) *When* I *is a* (p, ρ, γ) *Besov* r*-box for* f *about* $\mathbf{x}_0$*, there exists a positive* $r_0 < r$ *and* r_0*-box* J *about* $\mathbf{x}_0$ *such that* $\left\{\|f - \sigma_{2^n}(\mu_q^*; f)\|_{p,J}\right\} \in \mathsf{b}_{\rho,\gamma}$ *and* $\left\{\|\tau_n(\boldsymbol{\mu}_{\boldsymbol{q}}^*; f)\|_{p,J}\right\} \in \mathsf{b}_{\rho,\gamma}$.

(b) *When for some* r*-box* I *about* $\mathbf{x}_0$*, either* $\left\{\|f - \sigma_{2^n}(\mu_q^*; f)\|_{p,I}\right\} \in \mathsf{b}_{\rho,\gamma}$ *or* $\left\{\|\tau_n(\boldsymbol{\mu}_{\boldsymbol{q}}^*; f)\|_{p,I}\right\} \in \mathsf{b}_{\rho,\gamma}$*, then* I *is a* (p, ρ, γ) *Besov* r*-box for* f *about* $\mathbf{x}_0$.

Further, for $p = \infty$ *the same statements hold with* $\boldsymbol{\mu}$ *in place of* $\boldsymbol{\mu}_{\boldsymbol{q}}^*$ *and the elements* μ_n *in place of* μ_q^**.*

Again, these statements mirror those found in Theorem 2.23, but we prove them in a context for which the $p = \infty$ result may be easily obtained for use on scattered data.

Proof. Let $r > r_1 > r_0 > 0$ and suppose I is a (p, ρ, γ) Besov r-box for f about $\mathbf{x}_0$.

Let I_1, J be an r_1-box and r_0-box about $\mathbf{x}_0$, respectively. Let $\varphi \in C_I^\infty$ such that $\varphi \equiv 1$ on I_1. Then

$$\|f - \sigma_{2^n}(\mu_q^*; f)\|_{p,J} \leq \|f - f\varphi\|_{p,J} + \|f\varphi - \sigma_{2^n}(\mu_q^*; f\varphi)\|_{p,J} + \|\sigma_{2^n}(\mu_q^*; f\varphi - f)\|_{p,J} \quad .$$

The first term vanishes, as $f\varphi = f$ on $J \subset I_1$. The second term is bounded by its norm on the whole space, and so

$$\|f\varphi - \sigma_{2^n}(\mu_q^*; f\varphi)\|_{p,J} \leq \|f\varphi - \sigma_{2^n}(\mu_q^*; f\varphi)\|_p \leq cE_{2^{n-1},p}(f\varphi) \quad .$$

The third term satisfies the conditions of (3.2.10), and so

$$\|\sigma_{2^n}(\mu_q^*; f\varphi - f)\|_{p,J} \leq c\|f\varphi - f\|_1 2^{n(q-S)} \quad .$$

Since we have $2^{n\gamma} \cdot 2^{n(q-S)} = 2^{n(q+\gamma-S)}$ and $S > q+\gamma$, then $\sum |2^{n\rho(q+\gamma-S)}|$ converges and $\{2^{n(q-S)}\} \in \mathsf{b}_{\rho,\gamma}$. Since $f \in B^p_{\rho,\gamma}(\mathbf{x}_0)$ and we know φ is supported on a (p, ρ, γ) Besov r-box of f about $\mathbf{x}_0$, we conclude that $\{\|f - \sigma_{2^n}(\mu_q^*; f)\|_{p,J}\} \in \mathsf{b}_{\rho,\gamma}$. Further,

$$\|\tau_n(\boldsymbol{\mu}_{\boldsymbol{q}}^*; f)\|_{p,J} \leq \|f - \sigma_{2^n}(\mu_q^*; f)\|_{p,J} + \|f - \sigma_{2^{n-1}}(\mu_q^*; f)\|_{p,J} \quad ,$$

and so $\{\|\tau_n(\boldsymbol{\mu}_{\boldsymbol{q}}^*; f)\|_{p,J}\} \in \mathsf{b}_{\rho,\gamma}$ as well, yielding (a).

Now we will show (a) for the case of general measures. Let $p = \infty$. Then the proof proceeds very similarly, as

$$\|f - \sigma_{2^n}(\mu_n; f)\|_{\infty,J} \leq \|f - f\varphi\|_{\infty,J} + \|f\varphi - \sigma_{2^n}(\mu_n; f\varphi)\|_{\infty,J} + \|\sigma_{2^n}(\mu_n; f\varphi - f)\|_{\infty,J} \quad .$$

The first term is zero, the second is bounded by $cE_{2^{n-1},\infty}(f\varphi)$ due to (3.2.9) and the uniform boundedness of $|\!|\!|\mu_n|\!|\!|_{2^n}$, and the third is bounded by

$$c\|f\varphi - f\|_{\mu_n;1} 2^{n(q-S)} \leq c\|f\varphi - f\|_{\mu_n;\infty} 2^{n(q-S)} \leq c\|f\varphi - f\|_\infty 2^{n(q-S)} \quad ,$$

due to (3.2.10). Each term is the n^{th} entry of a sequence in $\mathsf{b}_{\rho,\gamma}$, and so $\{\|f-\sigma_{2^n}(\mu_n;f)\|_{\infty,J}\} \in \mathsf{b}_{\rho,\gamma}$. Likewise,

$$\|\tau_n(\boldsymbol{\mu};f)\|_{\infty,J} \le \|f-\sigma_{2^n}(\mu_n;f)\|_{\infty,J} + \|f-\sigma_{2^{n-1}}(\mu_{n-1};f)\|_{\infty,J} \quad .$$

Now we will show (b). Suppose I is an r-box about $\mathbf{x}_0$ for which $\big\{\|f-\sigma_{2^n}(\mu_q^*;f)\|_{p,I}\big\} \in \mathsf{b}_{\rho,\gamma}$. Let $\varphi \in C_I^\infty$. This means that there exists an $R \in \mathbb{H}_{2^n}^q$ with $\|\varphi - R\|_\infty \le c2^{-nS}$, as guaranteed by the direct theorem of approximation theory (see [4]). So, we may find

$$\begin{aligned} E_{2^{n+1},p}(f\varphi) &\le \|f\varphi - R\sigma_{2^n}(\mu_q^*;f)\|_p \\ &\le \|\varphi(f-\sigma_{2^n}(\mu_q^*;f))\|_p + \|(R-\varphi)\sigma_{2^n}(\mu_q^*;f)\|_p \\ &\le c(\varphi)\big(\|f-\sigma_{2^n}(\mu_q^*;f)\|_{p,I} + 2^{-nS}\|f\|_p\big) \quad . \end{aligned} \tag{3.2.15}$$

The first term which appears in (3.2.15) is the n^{th} element of a sequence lying in $\mathsf{b}_{\rho,\gamma}$ by hypothesis, and again as $S > \gamma$, the second term is as well.

Hence I is a (p,ρ,γ) Besov r-box for f about $\mathbf{x}_0$ and so $f \in B^p_{\rho,\gamma}(\mathbf{x}_0)$. If instead we assume I is an r-box about $\mathbf{x}_0$ for which $\Big\{\|\tau_n(\boldsymbol{\mu}_{\boldsymbol{q}}^*;f)\|_{p,I}\Big\} \in \mathsf{b}_{\rho,\gamma}$, the Hardy inequality (3.2.13) gives $\Big\{\|\tau_n(\boldsymbol{\mu}_{\boldsymbol{q}}^*;f)\|_{p,I}\Big\} \in \mathsf{b}_{\rho,\gamma} \Longrightarrow \big\{\|f-\sigma_{2^n}(\mu_q^*;f)\|_{p,I}\big\} \in \mathsf{b}_{\rho,\gamma}$, completing the proof of (b).

When we have $p=\infty$ for general measures, choosing R as before gives

$$\begin{aligned} E_{2^{n+1},\infty}(f\varphi) &\le \|f\varphi - R\sigma_{2^n}(\mu_n;f)\|_\infty \\ &\le \|\varphi(f-\sigma_{2^n}(\mu_n;f))\|_\infty + \|(R-\varphi)\sigma_{2^n}(\mu_n;f)\|_\infty \\ &\le c(\varphi)\big(\|f-\sigma_{2^n}(\mu_n;f)\|_{\infty,I} + 2^{-nS}\|f\|_{\mu_n;\infty}\big) \quad , \end{aligned} \tag{3.2.16}$$

the last step coming from the uniform boundedess of $\|\mu_n\|_{2^n}$ and from (3.2.8). Hence we reach the conclusion, $\{\|f - \sigma_{2^n}(\mu_n; f)\|_{\infty,I}\} \in \mathsf{b}_{\rho,\gamma}$, $\varphi \in C_I^\infty \Longrightarrow \{E_{2^n,\infty}(f\varphi)\} \in \mathsf{b}_{\rho,\gamma}$ and so I is a (∞, ρ, γ) Besov r-box for f about $\mathbf{x}_0$. The result using $\tau_n(\boldsymbol{\mu}; f)$ follows immediately by using the Hardy inequality (3.2.13) as before. □

3.3 PTN Construction and Approximation Properties

Having performed the same classification in the general M-Z case as was done in the Lebesgue case for multivariate trigonometric polynomials, we would like to extend the same results regarding local smoothness to PTN representations. The cornerstone of our analysis in the context of PTNs will be the opening theorem of the section. Here ν represents the desired centers of the PTN $\mathcal{I}_\phi(\nu; \mathcal{D}_\phi T)$.

Theorem 3.17. *For positive integers n, N, let $T \in \mathbb{H}_n^q$ and ν be an M-Z quadrature measure of order $n+N$. Let $\phi : \mathbb{T}^q \to \mathbb{C}$ have full frequency. Then for any $p \in [1, \infty]$,*

$$\|T - \mathcal{I}_\phi(\nu; \mathcal{D}_\phi T)\|_\infty \le c\|\nu\|_{n+N} \frac{E_{N,\infty}(\phi)}{m_n} n^{q\left(\frac{1}{p}-\frac{1}{2}\right)_+} \|T\|_p \quad .$$

Proving this theorem requires some preparation. First, we need the Riesz-Thorin interpolation theorem, so we reproduce a simplified version here.

Lemma 3.18. *Let $1 \le p_0, r_0, p_1, r_1 \le \infty$, and let F be a linear operator defined for simple functions on $\mathbb{T}^q$ such that*

$$\|F(f)\|_{r_j} \le M_j \|f\|_{p_j}, \qquad j = 0, 1$$

for all simple functions f. Let $0 < \theta < 1$, and

$$1/p = \theta/p_0 + (1-\theta)/p_1, \qquad 1/r = \theta/r_0 + (1-\theta)/r_1 \quad .$$

Then F is defined on L^p and we have

$$\|F(f)\|_r \leq M_0^{\theta} M_1^{(1-\theta)} \|f\|_p.$$

This is a direct consequence of the theorem as found in [29], for operators $\mathbb{T}^q \to \mathbb{T}^q$ equipped with the Lebesgue measure.

Additionally, we need the so–called Nikolskii inequalities (cf. [24]), for which we reproduce a similar proof to that found in ([15]).

Lemma 3.19. *Let $1 \leq p < r \leq \infty$, and let $P \in \mathbb{H}_n^q$. Then for some constant c,*

$$\|P\|_r \leq c n^{\left(\frac{1}{p}-\frac{1}{r}\right)q} \|P\|_p \quad . \tag{3.3.17}$$

Proof. In the case $r = \infty$, note that for all $\mathbf{x} \in \mathbb{T}^q$,

$$\|P\|_\infty^2 \leq |P(\mathbf{x})|^2 = \left| \sum_{|\mathbf{k}|_2 \leq n} \widehat{P}(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{x}} \right|^2 \quad .$$

Then using the Schwarz inequality (2.1.4) we can achieve the estimate

$$\|P\|_\infty^2 \leq \left(\sum_{|\mathbf{k}|_2 \leq n} \left| \widehat{P}(\mathbf{k}) \right|^2 \right) \left(\sum_{|\mathbf{k}|_2 \leq n} \left| e^{-i\mathbf{k}\cdot\mathbf{x}} \right|^2 \right) \leq c n^q \|P\|_2^2 \quad ,$$

where n^q is proportional to the cardinality of $\{\mathbf{k} \in \mathbb{Z}^q : |\mathbf{k}|_2 \leq n\}$; this is the volume of the Euclidean q-sphere.

We may also express $\|P\|_2^2$ as the integral $\int_{\mathbb{T}^q} |P(t)|^2 \mathrm{d}\mu_q^*(t)$ and apply the Hölder inequality (2.1.3) to achieve

$$\|P\|_2^2 = \int_{\mathbb{T}^q} |P(t)|^2 \mathrm{d}\mu_q^*(t) \leq \|P\|_\infty \|P\|_1 \leq c n^{q/2} \|P\|_2 \|P\|_1 \quad ,$$

which means $\|P\|_2 \leq c n^{q/2} \|P\|_1$, and so $\|P\|_\infty \leq c n^q \|P\|_1$.

If f is a simple function, we use this estimate with $2n$ in place of n, and $\sigma^*_{2n}(f)$ in place of P, for some S-smooth low-pass filter h with $S > q$. This gives (using (2.3.17) from Ch. 2)

$$\|\sigma^*_{2n}(f)\|_\infty \leq cn^q \|\sigma^*_{2n}(f)\|_1 \leq cn^q \|f\|_1 \quad .$$

Since $\|\sigma^*_{2n}(f)\|_\infty \leq c\|f\|_\infty$ as well, an application of the Riesz-Thorin interpolation theorem with $\theta = 1/p$ yields for $f \in L^p$, that

$$\|\sigma^*_{2n}(f)\|_\infty \leq cn^{q/p} \|f\|_p \quad .$$

Since $\|\sigma^*_{2n}(f)\|_p \leq c\|f\|_p$, one further application of Riesz-Thorin with $\theta = 1 - p/r$ yields for $f \in L^p$, that

$$\|\sigma^*_{2n}(f)\|_r \leq cn^{q(1/p-1/r)} \|f\|_p \quad .$$

Now, recall that $P \in \mathbb{H}^q_n \implies \sigma^*_{2n}(P) = P$. This completes the proof of (3.3.17). As a reminder, the constant c is arbitrary even up to different appearances in the same equation. $\square$

We are now ready to prove Theorem 3.17.

Proof. Choose $R \in \mathbb{H}^q_N$ with $\|\phi - R\|_\infty \leq 2E_{N,\infty}(\phi)$. Then, in this proof emphasizing reliance on μ^*_q,

$$\begin{aligned} \mathcal{I}_\phi(\mu^*_q; \mathcal{D}_\phi T, \mathbf{x}) - \mathcal{I}_\phi(\nu; \mathcal{D}_\phi T, \mathbf{x}) = \;& \mathcal{I}_{\phi-R}(\mu^*_q; \mathcal{D}_\phi T, \mathbf{x}) - \mathcal{I}_{\phi-R}(\nu; \mathcal{D}_\phi T, \mathbf{x}) \\ & + \mathcal{I}_R(\mu^*_q; \mathcal{D}_\phi T, \mathbf{x}) - \mathcal{I}_R(\nu; \mathcal{D}_\phi T, \mathbf{x}) \quad , \end{aligned}$$

by linearity of the integral. As $R(\mathbf{x} - \circ)\mathcal{D}_\phi T(\circ) \in \mathbb{H}^q_{n+N}$ and ν is a quadrature measure of order $n + N$, the last two terms are equal. As ν is an M-Z measure of order $n+N>n$, we use the general Young inequality in the form (3.2.12) to achieve

$$\begin{aligned} \|T-\mathcal{I}_\phi(\nu;\mathcal{D}_\phi T)\|_\infty &\leq \|\mathcal{I}_{\phi-R}(\mu_q^*;\mathcal{D}_\phi T,\mathbf{x})\|_\infty + \|\mathcal{I}_{\phi-R}(\nu;\mathcal{D}_\phi T,\mathbf{x})\|_\infty \\ &\leq c|\!|\!|\nu|\!|\!|_{n+N} E_{N,\infty}(\phi)\|\mathcal{D}_\phi T\|_1 \quad . \end{aligned} \tag{3.3.18}$$

Recalling the notation $m_n = \min_{|\mathbf{k}|_2\leq n} |\widehat{\phi}(\mathbf{k})|$, the ordering of the L^p spaces and the Parseval identity yield

$$\|\mathcal{D}_\phi T\|_1^2 \leq \|\mathcal{D}_\phi T\|_2^2 = \sum_{\ell\in\mathbb{Z}} |\widehat{\mathcal{D}_\phi T}(\ell)|^2 = \sum_{\ell\in\mathbb{Z}} \frac{|\widehat{T}(\ell)|^2}{|\widehat{\phi}(\ell)|^2} \leq \frac{1}{m_n^2}\sum_{|\ell|\leq n} |\widehat{T}(\ell)|^2 = \frac{1}{m_n^2}\|T\|_2^2 \quad , \tag{3.3.19}$$

noting that the sum (and hence m_n) is finite since $T\in\mathbb{H}_n$.

Now for $p\geq 2$, $\|T\|_2\leq\|T\|_p$ and we could show the result. For $p<2$, we must enlist the Nikolskii inequality, and so these cases collectively yield

$$\|T\|_2 \leq c n^{q\left(\frac{1}{p}-\frac{1}{2}\right)_+}\|T\|_p \quad . \tag{3.3.20}$$

Combining (3.3.20) with (3.3.18) and (3.3.19) shows the result

$$\|T-\mathcal{I}_\phi(\nu;\mathcal{D}_\phi T)\|_\infty \leq c|\!|\!|\nu|\!|\!|_{n+N} E_{N,\infty}(\phi)\|\mathcal{D}_\phi T\|_1 \leq c|\!|\!|\nu|\!|\!|_{n+N}\frac{E_{N,\infty}(\phi)}{m_n} n^{q\left(\frac{1}{p}-\frac{1}{2}\right)_+}\|T\|_p \quad .$$

□

We will use Theorem 3.17 to prove approximation results for arbitrary functions in X^p. Naturally, this approximation will be estimated by the degree of approximation of the target function by a suitable trigonometric polynomial, and that of the trigonometric polynomial by PTNs. To ensure that this second approximation is small enough not to destroy membership in the Besov spaces, we need to restrict the choice of ϕ. Observe that for some low-pass filter h,

$$E_{N,\infty}(\phi) \leq \|\phi - \sigma_N^*(h,\phi)\|_\infty \leq \sum_{|\mathbf{k}|_2 > N} |\widehat{\phi}(\mathbf{k})| \quad .$$

Definition 3.20. *Let $\beta > 0$. A function $\phi \in L^1$ will be said to have* ***β-receding frequency*** *when it has full frequency and*

$$\limsup_{n\to\infty} \left(\sum_{|\mathbf{k}|_2 > \beta n} \frac{|\widehat{\phi}(\mathbf{k})|}{m_n} \right)^{1/n} < 1 \quad .$$

Many standard examples exist of periodic functions for which the receding frequency criterion is satisfied, and which are used in network expansions. In each of the following examples, (2.3.10) is used together with results in [28] to deduce the stated Fourier coefficients.

EXAMPLE 3.3.1. Periodization of the Gaussian.

$$\phi(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^q} \exp(-|\mathbf{x} - 2\pi\mathbf{k}|_2^2/2) \ , \widehat{\phi}(\mathbf{k}) = (2\pi)^{q/2} \exp(-|\mathbf{k}|_2^2/2) \quad .$$

□

EXAMPLE 3.3.2. Periodization of the Hardy multiquadric.

$$\phi(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^q} (\alpha^2 + |\mathbf{x} - 2\pi\mathbf{k}|_2^2)^{-1} \ , \widehat{\phi}(\mathbf{k}) = \frac{\pi^{(q+1)/2}}{\Gamma\left(\frac{q+1}{2}\right)\alpha} \exp(-\alpha|\mathbf{k}|_2)$$

□

EXAMPLE 3.3.3. Tensor product construction using the Poisson kernel.

$$\phi(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{Z}^q} r^{-|\mathbf{k}|_1} e^{i\mathbf{k}\cdot\mathbf{x}} = \prod_{j=1}^{q} \frac{1 - r^2}{1 + r^2 - 2rx_j} \ , 0 < r < 1 \quad .$$

□

If we wish to approximate $f \in X^p$ with a PTN expansion, we should choose an appropriate ϕ with receding frequency, and then discretize $\mathcal{I}_\phi \mathcal{D}_\phi \sigma_n^*(f)$. To allow for the input of data other than the partial Fourier expanision, instead encode $\{\mu_n\}$ with a sequence of known information of f such that each measure is an M-Z quadrature measure, and use $\sigma_n(\mu_n; f)$.

Definition 3.21. *Let $\boldsymbol{\mu}, \boldsymbol{\nu}$ be sequences of Borel measures, both with nested support, and each ν_n additionally finitely supported. Let $\phi : \mathbb{T}^q \to \mathbb{C}$ have full frequency, and $h : \mathbb{R} \to \mathbb{R}$ be a low-pass filter (retained in the notation of this definition, for clarity). Then for $p \in [1, \infty]$, $f \in X^p$, the* ***PTN summability operator*** *is defined for each positive integer n by*

$$\mathcal{I}_\phi(\nu_n; \mathcal{D}_\phi \sigma_n(\mu_n; h, f))(\mathbf{x}) = \int_{\mathbb{T}^q} \phi(\mathbf{x} - \mathbf{y}) \; \mathcal{D}_\phi \left[\sigma_n(\mu_n; h, f)\right](\mathbf{y}) \; \mathrm{d}\nu_n(\mathbf{y}) \quad .$$

Notice that as each ν_n has finite support, each such operator is a PTN as desired, since it resides in the span of $\{\phi(\mathbf{x} - \mathbf{z}) | \; \mathbf{z} \in \mathrm{supp}(\nu_n)\}$.

We take a moment to emphasize that when the elements of the measure sequence $\boldsymbol{\mu}$ with respect to which $\sigma_n(\mu_n; h, f)$ operates on f are all taken to be the Lebesgue measure, the operator produces a linear combination of Fourier coefficients for each n. When the measures are finitely supported, the operator produces linear combination of function values for each n. More generality is possible here too, within the most general theory of M-Z measures and quadrature measures - we will not discuss the subject, but it may be pursued in any of [6, 11, 19], to name only some relevant papers studied in the preparation of this thesis.

As we will see, these PTN operators approximate f well so as long as trigonometric polynomials approximate ϕ sufficiently well - specifically we will require that ϕ has β-receding frequency so that this degree of approximation is at least exponential, placing the PTN operator in the same Besov space as $\sigma_n(f)$. Then the desired approximation follows from the results of Chapter 2 and Section 3.2. This proof follows the method found in [19]. We recall again that $|\!|\!|\mu|\!|\!|_n \sim |\!|\!|\mu|\!|\!|_{cn}$.

Proposition 3.22. *Let β be a positive integer, and let $\phi : \mathbb{T}^q \to \mathbb{C}$ have β-receding frequency. Let $\tilde{\boldsymbol{\mu}}$ be a sequence of M-Z quadrature measures with nested support, each of which has order $3/2\ n$, with uniformly bounded M-Z norm $|\!|\!|\tilde{\mu}_n|\!|\!|_n \leq c$, for all n, independent of n. Let $\tilde{\boldsymbol{\nu}}$ be a sequence of finitely supported M-Z quadrature measures with nested support, each of which has order $(1+\beta)n$, with uniformly bounded M-Z norm $|\!|\!|\tilde{\nu}_n|\!|\!|_n \leq c$, for all n, independent of n. Then if for some $p \in [1,\infty]$, $f \in X^p$, it is the case for each positive integer n that*

$$\|\sigma_n(\mu_q^*; f) - \mathcal{I}_\phi(\tilde{\nu}_n; \mathcal{D}_\phi \sigma_n(\mu_q^*; f))\|_p \leq c \frac{E_{\beta n,\infty}(\phi)}{m_n} n^{q\left(\frac{1}{p}-\frac{1}{2}\right)_+} \|f\|_p \tag{3.3.21}$$

and

$$\|\sigma_n(\tilde{\mu}_n; f) - \mathcal{I}_\phi(\tilde{\nu}_n; \mathcal{D}_\phi \sigma_n(\tilde{\mu}_n; f))\|_\infty \leq c \frac{E_{\beta n,\infty}(\phi)}{m_n} \|f\|_\infty \quad . \tag{3.3.22}$$

Proof. Note that $\sigma_n(\mu_q^*; f) \in \mathbb{H}_n^q$ and ϕ has β-receding frequency. So, the conditions of Theorem 3.17 are met and we may choose $N = \beta n$. Hence, we obtain

$$\begin{aligned}
\|\mathcal{I}_\phi(\tilde{\nu}_n; \mathcal{D}_\phi \sigma_n(\mu_q^*; f)) - \sigma_n(\mu_q^*; f)\|_p &\leq \|\mathcal{I}_\phi(\tilde{\nu}_n; \mathcal{D}_\phi \sigma_n(\mu_q^*; f)) - \sigma_n(\mu_q^*; f)\|_\infty \\
&\leq c|\!|\!|\tilde{\nu}_n|\!|\!|_n \frac{E_{\beta n,\infty}(\phi)}{m_n} n^{q\left(\frac{1}{p}-\frac{1}{2}\right)_+} \|f\|_p \\
&\leq c \frac{E_{\beta n,\infty}(\phi)}{m_n} n^{q\left(\frac{1}{p}-\frac{1}{2}\right)_+} \|f\|_p \quad ,
\end{aligned}$$

enlisting (3.2.8) and the uniform boundedness of $\|\tilde{\nu}_n\|_n$. When $p = \infty$, we have for the general measure case,

$$\begin{aligned}
\|\mathcal{I}_\phi(\tilde{\nu}_n; \mathcal{D}_\phi \sigma_n(\tilde{\mu}_n; f)) - \sigma_n(\tilde{\mu}_n; f)\|_\infty &\leq c\|\tilde{\nu}_n\|_n \frac{E_{\beta n,\infty}(\phi)}{m_n} \|\sigma_n(\tilde{\mu}_n; f)\|_\infty \\
&\leq c\|\tilde{\nu}_n\|_n \|\tilde{\mu}_n\|_n \frac{E_{\beta n,\infty}(\phi)}{m_n} \|f\|_\infty \\
&\leq c\frac{E_{\beta n,\infty}(\phi)}{m_n} \|f\|_\infty \quad ,
\end{aligned}$$

using uniform boundedness of both measure sequences on the last step. Note that $p = \infty > 2$ means the exponent of the n^q term of Theorem 3.17 vanishes. □

It is convenient here to express the property of β-receding frequency in the form

$$\frac{E_{\beta n,\infty}(\phi)}{m_n} \leq c_0 2^{-bn} \quad , \tag{3.3.23}$$

for some positive constants c_0, b. Notice that the same then holds for $\frac{E_{\beta n,\infty}(\phi)}{m_n} n^{c_1 q}$ with a different choice of c_0 and b, for any c_1, for sufficiently large n. Lastly, it is clear that

$$\|f - \mathcal{I}_\phi(\tilde{\nu}_n; \mathcal{D}_\phi \sigma_n(\mu_q^*; f))\|_p \leq \|f - \sigma_n(\mu_q^*; f)\|_p + \|\sigma_n(\mu_q^*; f) - \mathcal{I}_\phi(\tilde{\nu}_n; \mathcal{D}_\phi \sigma_n(\mu_q^*; f))\|_p \quad ,$$

and

$$\|f - \mathcal{I}_\phi(\tilde{\nu}_n; \mathcal{D}_\phi \sigma_n(\tilde{\mu}_n; f))\|_\infty \leq \|f - \sigma_n(\tilde{\mu}_n; f)\|_\infty + \|\sigma_n(\tilde{\mu}_n; f) - \mathcal{I}_\phi(\tilde{\nu}_n; \mathcal{D}_\phi \sigma_n(\tilde{\mu}_n; f))\|_\infty \quad .$$

We pause to develop a set of conditions and related shorthand which will form the setting for the remaining results of this chapter, as we construct, and elucidate the properties of the wavelet-like PTN expansion. Note the change from the conditions of Proposition 3.22 to use measure sequences whose elements are M-Z measures of dyadic order - this change simplifies the proofs regarding Besov spaces considerably, and in practice one may form from any sufficiently dense data, the nested supports of a desired number of elements from such a sequence (see [14]).

Remark 3.23. *Select a positive integer β, and a function $\phi : \mathbb{T}^q \to \mathbb{C}$ having β-receding frequency. Let $\boldsymbol{\mu}$ be a sequence of M-Z quadrature measures with nested support, each of which has order $3 \cdot 2^{n-1}$, with uniformly bounded M-Z norm $|||\mu_n|||_{2^n} \leq c$, for all n, independent of n. Let $\boldsymbol{\nu}$ be a sequence of finitely supported M-Z quadrature measures with nested support, each of which has order $(1+\beta)2^n$, with uniformly bounded M-Z norm $|||\nu_n|||_{2^n} \leq c$, for all n, independent of n.*

When for some $p \in [1, \infty]$, $f \in X^p$, we denote for $\mathbf{x} \in \mathbb{T}^q$,

$$G_n^*(f, \mathbf{x}) := \mathcal{I}_\phi(\nu_n; \mathcal{D}_\phi \sigma_{2^n}(\mu_q^*; f))(\mathbf{x}) \quad .$$

When $f \in X^\infty$, we denote for $\mathbf{x} \in \mathbb{T}^q$,

$$G_n(f, \mathbf{x}) := \mathcal{I}_\phi(\nu_n; \mathcal{D}_\phi \sigma_{2^n}(\mu_n; f))(\mathbf{x}) \quad .$$

Summarizing the above results and statements, we enlist this notation and use (3.3.23) to obtain a corollary of Proposition 3.22.

Corollary 3.24. *Assume the conditions of Remark* 3.23*. Then if for some $p \in [1, \infty]$, $f \in X^p$, we have*

$$\|\sigma_{2^n}(\mu_q^*; f) - G_n^*(f)\|_p \leq c 2^{-b2^n} \|f\|_p \quad , \tag{3.3.24}$$

and

$$\|f - G_n^*(f)\|_p \leq c \left(E_{2^{n-1},p}(f) + 2^{-b2^n} \|f\|_p \right) \quad . \tag{3.3.25}$$

Additionally, when $p = \infty$ and $f \in X^\infty$, we have

$$\|\sigma_{2^n}(\mu_n; f) - G_n(f)\|_\infty \le c2^{-b2^n} \|f\|_\infty \quad , \tag{3.3.26}$$

and

$$\|f - G_n(f)\|_\infty \le c\left(E_{2^{n-1},\infty}(f) + 2^{-b2^n}\|f\|_\infty\right) \quad . \tag{3.3.27}$$

Note that the constants in each equation are distinct from, and unrelated to the constants in any other equation.

3.4 Classification of Smoothness and Local Smoothness

In a manner analogous to that of Chapter 2 and Section 3.2, we conclude that the G_n summability operators create a sufficiently localized partial expansion of f for corresponding band-pass operators to behave as a wavelet-like expansion. We define those band-pass operators now.

Definition 3.25. *Assume the conditions of Remark* 3.23. *Then for $p \in [1, \infty]$, $f \in X^p$, the* ***PTN band-pass operator*** *is defined for each non-negative integer n by*

$$\begin{aligned}
\mathcal{T}_0(\boldsymbol{\nu}, \boldsymbol{\mu}; h, \phi, f)(\mathbf{x}) &:= \mathcal{I}_\phi(\nu_0; \mathcal{D}_\phi\sigma_1(\mu_0; h, f))(\mathbf{x}) = G_0(f), \qquad \textit{and for } n \ge 1 \quad , \\
\mathcal{T}_n(\boldsymbol{\nu}, \boldsymbol{\mu}; h, \phi, f)(\mathbf{x}) &:= \mathcal{I}_\phi(\nu_n; \mathcal{D}_\phi\sigma_{2^n}(\mu_n; h, f))(\mathbf{x}) - \mathcal{I}_\phi(\nu_{n-1}; \mathcal{D}_\phi\sigma_{2^{n-1}}(\mu_{n-1}; h, f))(\mathbf{x}) \\
&= G_n(f, \mathbf{x}) - G_{n-1}(f, \mathbf{x}) \quad .
\end{aligned}$$

We will immediately use the shorthand

$$\mathcal{T}_n(f, \mathbf{x}) := \mathcal{T}_n(\boldsymbol{\nu}, \boldsymbol{\mu}; h, \phi, f)(\mathbf{x}) \quad ,$$

whenever we assume the conditions of Remark 3.23 and the assignment of each omitted notation is clear. Similarly, for when we omit the case of scattered data and require $\boldsymbol{\mu} = \boldsymbol{\mu}^*_{\boldsymbol{q}}$, we define

$$\mathcal{T}^*_0(f,\mathbf{x}) := G^*_0(f,\mathbf{x}) \ , \ \text{and for } n \geq 1, \ \mathcal{T}^*_n(f,\mathbf{x}) := G^*_n(f,\mathbf{x}) - G^*_{n-1}(f,\mathbf{x}) \ .$$

Proposition 3.26. *Assume the conditions of Remark* 3.23*. Then if for some* $p \in [1,\infty]$*,* $f \in X^p$*, we have*

$$f(\mathbf{x}) = \sum_{j=0}^{\infty} \mathcal{T}^*_j(f,\mathbf{x}) \quad \textit{and when } p = \infty, \textit{ we have} \quad f(\mathbf{x}) = \sum_{j=0}^{\infty} \mathcal{T}_j(f,\mathbf{x}) \ .$$

Convergence of the series is in the sense of L^p *and of* L^∞*, respectively.*

Proof. The partial sums of PTN band-pass operators form a telescoping series, and so by (3.3.25),

$$\left\| f - \sum_{j=0}^{n} \mathcal{T}^*_j(f) \right\|_p = \|f - G^*_n(f)\|_p \leq c\left(E_{2^{n-1},p}(f) + 2^{-b2^n}\|f\|_p\right) \to 0 \ \text{ as } \ n \to \infty \ .$$

Likewise, when $p = \infty$, we enlist (3.3.27) to achieve,

$$\left\| f - \sum_{j=0}^{n} \mathcal{T}_j(f) \right\|_\infty = \|f - G_n(f)\|_\infty \leq c\left(E_{2^{n-1},\infty}(f) + 2^{-b2^n}\|f\|_\infty\right) \to 0 \ \text{ as } \ n \to \infty \ .$$

□

The form of Corollary 3.24 now enables us to very concisely state and prove the Besov space and local Besov space characterizations for the PTN network operators, completing the characterization of (local) smoothness. Note that for $\rho, \gamma, b > 0$, $\sum_{n=0}^{\infty} 2^{\rho(n\gamma - b2^n)}$ is convergent by the Cauchy root test, and $2^{n\gamma - b2^n}$ is bounded. Hence

$\{2^{-b2^n}\} \in \mathsf{b}_{\rho,\gamma}$ for any choice of ρ, γ. Summarily, under the conditions of Remark 3.23, given $\gamma > 0$, $\rho \in (0, \infty]$, then for $p \in [1, \infty]$ and $f \in X^p$,

$$\{\|f - \sigma_{2^n}(\mu_q^*; f)\|_p\} \in \mathsf{b}_{\rho,\gamma} \iff \{\|f - G_n^*(f)\|_p\} \in \mathsf{b}_{\rho,\gamma} \quad , \tag{3.4.28}$$

and for $f \in X^\infty$,

$$\{\|f - \sigma_{2^n}(\mu_n; f)\|_\infty\} \in \mathsf{b}_{\rho,\gamma} \iff \{\|f - G_n(f)\|_\infty\} \in \mathsf{b}_{\rho,\gamma} \quad . \tag{3.4.29}$$

The following characterization theorems are now mainly corollaries of their trigonometric counterparts.

Theorem 3.27. *Assume the conditions of Remark* 3.23*. Then for* $\gamma > 0$, $\rho \in (0, \infty]$, $p \in [1, \infty]$*, and* $f \in X^p$*, the following statements are equivalent:*

(a) $f \in B^p_{\rho,\gamma}$*, and so* $\{E_{2^n,p}(f)\} \in \mathsf{b}_{\rho,\gamma}$.

(b) $\{\|f - G_n^*(f)\|_p\} \in \mathsf{b}_{\rho,\gamma}$

(c) $\{\|\mathcal{T}_n^*(f)\|_p\} \in \mathsf{b}_{\rho,\gamma}$

Additionally, for $p = \infty$ *and* $f \in X^p$*, the same equivalences hold with* $G_n(f)$ *in place of* $G_n^*(f)$ *and* $\mathcal{T}_n(f)$ *in place of* $\mathcal{T}_n^*(f)$*.*

Proof. The only equivalence which does not come immediately from Theorem 3.14, (3.4.28), and (3.4.29) is (b) $\iff$ (c), and this may be shown as in the trigonometric polynomial case. First notice that $\|\mathcal{T}_n^*(f)\|_p \le \|f - G_n^*(f)\|_p + \|f - G_{n-1}^*(f)\|_p$ to achieve one direction. Then using Proposition 3.26, we see that

$$\|f - G_n^*(f)\|_p \le \left\| \sum_{j=0}^{\infty} \mathcal{T}_j^*(f) - \sum_{j=0}^{n} \mathcal{T}_j^*(f) \right\|_p \le \sum_{j=n+1}^{\infty} \|\mathcal{T}_j^*(f)\|_p \quad .$$

The Hardy inequality (3.2.13) with $\upsilon = 1$ then yields the reverse direction. As Proposition 3.26 accounts for the band-pass operators with respect to M-Z measures, when $p = \infty$ we achieve the same equivalence with $G_n(f)$ and $\mathcal{T}_n(f)$ in place of $G_n^*(f)$ and $\mathcal{T}_n^*(f)$, respectively. $\square$

In this way, the PTN operators can achieve the same characterization of a target function's smoothness, in terms of Besov spaces. What remains is to further characterize local smoothness in terms of local Besov spaces, and we show that PTNs can accomplish this as well.

Theorem 3.28. *Assume the conditions of Remark* 3.23*, and additionally let* $S > q + \gamma$*, and let* $\mathbf{x}_0 \in \mathbb{T}^q$*. Let* $\gamma > 0$*,* $\rho \in (0, \infty]$*,* $p \in [1, \infty]$*, and* $f \in X^p$*. Then we have the following:*

(a) *When* I *is a* (p, ρ, γ) *Besov* r*-box for* f *about* $\mathbf{x}_0$*, there exists a positive* $r_0 < r$ *and an* r_0*-box* J *about* $\mathbf{x}_0$ *such that* $\{\|f - G_n^*(f)\|_{p,J}\} \in \mathsf{b}_{\rho,\gamma}$ *and* $\{\|\mathcal{T}_n^*(f)\|_{p,J}\} \in \mathsf{b}_{\rho,\gamma}$.

(b) *When for some* r*-box* I *about* $\mathbf{x}_0$*, either* $\{\|f - G_n^*(f)\|_{p,I}\} \in \mathsf{b}_{\rho,\gamma}$ *or* $\{\|\mathcal{T}_n^*(f)\|_{p,I}\} \in \mathsf{b}_{\rho,\gamma}$*, then* I *is a* (p, ρ, γ) *Besov* r*-box for* f *about* $\mathbf{x}_0$.

Further, for $p = \infty$ *the same statements hold with* $G_n(f)$ *and* $\mathcal{T}_n(f)$ *in place of* $G_n^*(f)$ *and* $\mathcal{T}_n^*(f)$*, respectively.*

We rely on the prior result Theorem 3.16, to include the PTN results.

Proof. In order to utilize the exponential closeness of the PTN summability operators to their corresponding trigonometric polynomial operators in characterizing the local smoothness classes, we must bound their local norms. Let $r > 0$ and I be an r-box about $\mathbf{x}_0$. Then by (3.3.24),

$$\begin{aligned}\Big| \|f - \sigma_{2^n}(\mu_q^*; f)\|_{p,I} - \|f - G_n^*(f)\|_{p,I} \Big| &\leq \|\sigma_{2^n}(\mu_q^*; f) - G_n^*(f)\|_{p,I} \\ &\leq \|\sigma_{2^n}(\mu_q^*; f) - G_n^*(f)\|_p \leq c2^{-b2^n}, \end{aligned} \tag{3.4.30}$$

and by (3.3.26),

$$\begin{aligned}\Big| \|f - \sigma_{2^n}(\mu_n; f)\|_{p,I} - \|f - G_n(f)\|_{p,I} \Big| &\leq \|\sigma_{2^n}(\mu_n; f) - G_n(f)\|_{p,I} \\ &\leq \|\sigma_{2^n}(\mu_n; f) - G_n(f)\|_p \leq c2^{-b2^n}. \end{aligned} \tag{3.4.31}$$

Note that these sequences of differences lie in $\mathsf{b}_{\rho,\gamma}$. Then by this fact and Theorem 3.16 (a), we conclude that (a) holds, and additionally the same holds when $p = \infty$ and $G_n(f)$ and $\mathcal{T}_n(f)$ are used in place of $G_n^*(f)$ and $\mathcal{T}_n^*(f)$, respectively.

We have that (b) follows immediately as well from (3.4.30), since

$$\{\|\mathcal{T}_n^*(f)\|_{p,I}\} \in \mathsf{b}_{\rho,\gamma} \Longrightarrow \{\|f - G_n^*(f)\|_{p,I}\} \in \mathsf{b}_{\rho,\gamma} \Longrightarrow \left\{\|f - \sigma_{2^n}(\mu_q^*; f)\|_{p,I}\right\} \in \mathsf{b}_{\rho,\gamma} \quad ,$$

and by Theorem 3.16 (b) we have that $\left\{\|f - \sigma_{2^n}(\mu_q^*; f)\|_{p,I}\right\} \in \mathsf{b}_{\rho,\gamma} \Longrightarrow I$ is a (p, ρ, γ) Besov r-box for f about $\mathbf{x}_0$.

Further, when $p = \infty$, analogously using (3.4.31),

$$\{\|\mathcal{T}_n(f)\|_{\infty,I}\} \in \mathsf{b}_{\rho,\gamma} \Longrightarrow \{\|f - G_n(f)\|_{\infty,I}\} \in \mathsf{b}_{\rho,\gamma} \Longrightarrow \{\|f - \sigma_{2^n}(\mu_n; f)\|_{\infty,I}\} \in \mathsf{b}_{\rho,\gamma} \quad ,$$

and again by Theorem 3.16 (b) we conclude that (b) holds for the general measure case. □

References

[1] C. de Boor, *Polynomial Interpolation in Several Variables,* 1994, cached version available at http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.23.3079.

[2] A. Bornstein, J. Brand, M. McVey, and A. Neiderer, *Fusion of asynchronous, parallel, unreliable data streams,* ARL-TR-5344, September, 2010.

[3] I. Daubeschies, "Ten lectures on wavelets, CBMS-NSF Series in Appl. Math.", SIAM Publications, Philadelphia, 1992.

[4] R. A. DeVore and G. G. Lorentz, "Constructive approximation", Springer Verlag, Berlin, 1993.

[5] N. Dyn, *Interpolation and Approximation by Radial and Related Functions,* Approximation Theory VI: Vol. I (C. K. Chui, L. L. Schumaker, and J. D. Ward Eds.), Academic Press, pp. 211-234.

[6] F. Filbir and H. N. Mhaskar, *Marcinkiewicz–Zygmund measures on manifolds,* Journal of Complexity, **27** (2011), 568–596.

[7] F. Girosi, M. Jones and T. Poggio, *Regularization theory and neural networks architectures,* Neural Computation, **7** (1995), 219-269.

[8] B. V. Gnedenko, "Theory of probability", Gordon and Breach Science, Australia, 1997.

[9] D. Harrison and D. L. Rubinfeld, *Hedonic prices and the demand for clean air,* Journal of Environmental Economics and Management, **5** (1978), 81–102.

[10] L. Khachikyan and H. N. Mhaskar, *Neural networks for function approximation,* in "Neural networks for signal processing, V", (F. Girosi, J. Makhoul, E. Manolakos, E. Wilson Eds.), IEEE, New York, 1995, pp.21–29.

[11] Q. T. Le Gia and H. N. Mhaskar, *Localized linear polynomial operators and quadrature formulas on the sphere,* SIAM J. Numer. Anal. **47** (1) (2008), 440–466.

[12] M. Golomb, H. F. Weinberger, *Optimal approximation and error bounds,* On Numerical Approximation, (R.E. Langer Ed.), Univ. Wisconsin Press, Madison 1959, pp. 117-190.

[13] H. N. Mhaskar, *Approximation theory and neural networks,* in "Wavelet Analysis and Applications, Proceedings of the international workshop in Delhi, 1999", (P. K. Jain, M. Krishnan, H. N. Mhaskar, J. Prestin, and D. Singh Eds.), Narosa Publishing, New Delhi, India, 2001, 247–289.

[14] H. N. Mhaskar, *Eignets for function approximation,* Appl. Comput. Harmon. Anal. **29**, (2010), 63-87.

[15] H. N. Mhaskar, "Local analysis of spectral and scattered data", in preparation, to be publised by Atlantis Publications (Springer Verlag).

[16] H. N. Mhaskar, *Polynomial operators and local smoothness classes on the unit interval, II,* Jaén J. of Approx., **1** (1) (2009), 1–25.

[17] H. N. Mhaskar and C. A. Micchelli, *Approximation by superposition of a sigmoidal function and radial basis functions,* Advances in Applied Mathematics, **13** (1992), 350-373.

[18] H. N. Mhaskar and C. A. Micchelli, *Degree of approximation by neural and translation networks with a single hidden layer,* Advances in Applied Mathematics, **16** (1995), 151-183.

[19] H. N. Mhaskar, F. J. Narcowich, and J. D. Ward, *Spherical Marcinkiewicz-Zygmund inequalities and positive quadrature,* Math. Comp. 70 (2001), no. 235, 1113–1130.

[20] H. N. Mhaskar and D. V. Pai, "Fundamentals of Approximation Theory", Narosa Publishing Co., Delhi, 2000.

[21] H. N. Mhaskar and J. Prestin, *On the detection of singularities of a periodic function,* Advances in Computational Mathematics, **12** (2000), 95–131.

[22] H. N. Mhaskar and J. Prestin, *Polynomial frames: a fast tour,* in "Approximation Theory XI, Gatlinburg, 2004" (C. K. Chui, M. Neamtu, and L. Schumaker Eds.), Nashboro Press, Brentwood, 2005, 287–318.

[23] H. N. Mhaskar and J. Prestin, *On local smoothness classes of periodic functions,* Journal of Fourier Analysis and Applications, **11** (3) (2005), 353 - 373.

[24] S. M. Nikolskii, *Inequalities for entire functions of finite degree and their application in the theory of differentiable functions of several variables,* Trudy Mat. Inst. Steklov, **38** (1951), 244 – 278.

[25] P. Niyogi and F. Girosi, *On the relationship between generalized error, hypothesis complexity, and sample complexity for radial basis functions,* Neural Computation, **8** (1996), 819–842.

[26] W. Rudin, “Real and complex analysis”, McGraw Hill, New York, 1974.

[27] R. Schaback, *Native Hilbert spaces for radial basis functions. I,* New developments in approximation theory (Dortmund, 1998), Internat. Ser. Numer. Math., 132, Birkhäuser, Basel, 1999, pp. 255–282.

[28] E. M. Stein and G. Weiss, “Fourier Analysis on Euclidean Spaces”, Princeton University Press, Princeton, New Jersey, 1971.

[29] A. Zygmund, “Trigonometric Series”, Cambridge University Press, Cambridge, 1977.